Prevalence of the gene *lsr2* among the genus *Mycobacterium* and an investigation into changes in biofilm formation when inactivating *lsr2* regulated genes in *Mycobacterium smegmatis*

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Wayne C. Gatlin III

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

John L. Dahl

October, 2014

Acknowledgements

I would like to express my heartfelt thanks to my advisor Dr. John L. Dahl, whose guidance and sense of wonder invigorated my own scientific curiosities. Dr. Dahl showed me how science can be a multidisciplinary experience that feeds off all forms of the arts. I would also like to thank Dr. Dahl for showing me how to be a good teacher, of which he is the gold standard.

Abstract

There are many *Mycobacterium* species found throughout the world, some capable of causing human disease or industrial problems. *Mycobacterium tuberculosis* is arguably the most infamous, however, there are over 150 other species of *Mycobacterium* that have also been identified. Many *mycobacterium* are capable of forming biofilms, which are complex matrixes of bacterial cells that can adhere to surfaces. Recently, a strain of *Mycobacterium smegmatis* was characterized that lacked the ability to form biofilms. This phenotype has been linked to a mutation leading to the absence of the DNA bridging protein Lsr2. This study describes the further characterization of the role the *lsr2* gene plays in biofilm formation in *M. smegmatis* and the prevalence of the gene among other species of the genus *Mycobacterium.* This study reports that *lsr2* is found in 46 of 52 *Mycobacterial* species tested. The gene was also identified as a highly polymorphic site that could be used to identify *mycobacterial* species by DNA sequence analysis.

The role that *lsr2* plays in regulating *M. smegmatis* biofilm formation was investigated by inactivating genes hypothesized to be regulated by Lsr2. Genetic knockouts lacking both *lsr2* and *lsr2*-regulated genes were evaluated for their abilities to form biofilms, as compared to biofilm formation for wild-type *M. smegmatis.* We report a link between the *mas* (mycocerosic acid synthase) gene and the ability of the Lsr2 mutant strain of *M. smegmatis* to form a biofilm.

Table of Contents

List of Tables

List of Figures

A survey of 52 *Mycobacterial* species for the presence of the gene *lsr2*

ABSTRACT

Previous studies have shown that Lsr2 influences expression of biofilm formation in *Mycobacterium smegmatis*, yet Lsr2 function is largely unknown in other *Mycobacterium* species. The current study describes the prevalence of *lsr2* among type strains of 54 *Mycobacterium* species and assesses the significance of the gene as a sequenceable phylogenetic marker for the *Mycobacterium* genus. In this study a 325-bp internal fragment of the *lsr2* gene was amplified by PCR and sequenced for each of the 52 type strains tested. Presence of the gene was determined by *in silico* comparison of the sequenced PCR products to previously published *lsr2* sequences. Results show that 46 of the 52 species tested have sequences that show 74-100% homology to *lsr2* from published strains. Secondly, a phylogenetic analysis of the DNA sequences and published *lsr2* sequence from the NCBI database indicates clear nucleotide differences between many rapidly growing *Mycobacterium*. However, there is no difference between several members of the *M. tb* complex (*M. tuberculosis, M. canettii, M. africanum*), no difference between *M. marinum* and *M. ulcerans*, no difference between *M. senegalense* and *M. peregranum*, and no difference between *M. goodii* and *M. cookii*. Because of low sequence differences and small fragment size, *lsr2* is likely not a candidate gene for future phylogenetic analysis.

INTRODUCTION

There are many *Mycobacterium* species found throughout the world, some capable of

causing human disease or industrial problems. *Mycobacterium tuberculosis* is arguably the most infamous, however, there are over 150 other species of *Mycobacterium* that have also been identified. Recently, *Mycobacterium* species living in biofilms have attracted attention because these bacteria can grow in higher concentrations than planktonic cells and resist disinfectants (Hall-Stoodley, 2004)(Falkinham, 2001)(Feazel, 2009). Biofilms result when bacterial cells attach to a surface and begin to multiply. A matrix of cells with extracellular polymeric substances (EPS) can form over time and facilitate attachment of additional bacterial species that can also contribute to the heterogeneity of the biofilm (Stoodley, 2002). Many *Mycobacterium* species are known to form biofilms, including *Mycobacterium tuberculosis, Mycobacterium fortuitum,* and *Mycobacterium smegmatis* (Ojha, 2008)(Hall-Stoodley, 2006).

Recently a cytoplasmic protein called Lsr2 has been associated with the ability of *M. smegmatis* to form biofilms (Chen 2006). The Lsr2 protein was originally described as an antigen expressed by *Mycobacterium leprae* during infections (Laal, 1991), and recent studies have shown that the protein is responsible for regulating a diverse range of cellular processes (Colangeli, 2009)(Nguyen, 2010). One well-studied function of Lsr2 is its role as a histone-like DNA binding protein, where it can influence the expression of over 60 genes in the *M. smegmatis* genome (Colangeli, 2007).

Previous studies have shown that deactivating *lsr2* in *M. smegmatis* will cause a phenotypic change (Chen, 2006)(Arora, 2008). Notably, the bacteria lose their ability to form biofilms on plastic surfaces or floating pellicles in liquid media (Chen, 2006)(Arora,

2008). It has been hypothesized that this is due to alteration of surface molecules, such as mycolyl-diacylglycerols and other apolar lipids, when Lsr2 is not present (Chen, 2006).

Currently, only a few *Mycobacterium* species are known to contain *lsr2*, as evident by examining the few annotated genomes available. In this study I investigated the prevalence of *lsr2* in the *Mycobacterium* genus by searching for *lsr2* gene homologs in 52 well-characterized *Mycobacterium* species that are available in the public domain. Species-level identification of *Mycobacterium* using 16 S rRNA gene sequencing is not generally successful, and only a few genes are currently used as sequencable markers, therefore, I also investigated whether *lsr2* would make a suitable phylogenetic marker for identifying *Mycobacterium* to the species level.

MATERIALS AND METHODS

Bacterial strains and culture conditions

The 52 *Mycobacterium* strains and 8 published sequences that were used in this study are listed in Table 1. All strains were cultured on Middlebrook 7H11 medium (Difco) supplemented with OADC (Sigma-Aldrich) and incubated at either 30˚C or 37˚C, as per ATCC recommendations.

Primer Design

Two conserved regions were identified within the *lsr2* gene after aligning sequences from four *Mycobacterium* species which were available online in the National Center for Biotechnology Information (NCBI) GenBank database. These regions were used to

design two primers (lsr2F and lsr2R), which correspond to *M. smegmatis lsr2* gene positions 1 to 23 and 302 to 325, respectively. Forward primer lsr2F (ATG GCK AAG AAA GTN ACC GTC AC) and reverse primer lsr2R (ATG ACG TCG GCC GGG ATS CGG CC) were designed to universally amplify an internal DNA fragment of *lsr2* (referred to as '*lsr2*') from *Mycobacterium* species. These primers contain degenerate bases at locations K, N, and S that are variable in a multiple species alignment of published *lsr2* genes.

DNA Extraction

Mycobacterium samples were swabbed from plated cultures to obtain DNA for PCR amplification. Samples were physically lysed to extract template DNA by "bead beating". Briefly, a swab of bacterial sample was transferred to an impact-resistant screw-capped vessel containing PBS and 0.1 mm diameter glass beads and homogenized using a Thermo Savant FastPrep FP120 Homogenizer. Cell debris was removed by brief centrifugation and remaining suspension was removed and quantified for nucleic acid by Nanodrop spectrophotometery.

PCR amplification and Sequencing

PCR amplification was performed using Illustra Puretaq Ready-To-Go PCR beads (GE Healthcare) in 0.2-mL reaction tubes, according to the protocol of the supplier. DMSO was added to reduce the formation of secondary structures in ssDNA. Each reaction contained dH_2O, 100 mM lsr2F primer, 100 mM lsr2R primer, 2% DMSO (v/v), and 2 µL template DNA.

The thermal profile involved an initial denaturation of 95˚C for 5 min followed by 35 cycles of 95˚C for 30 sec, 59˚C for 30 sec, and 72˚C for 1 min. A final 5-min 72˚C elongation and subsequent 4˚C holding completed the reaction.

PCR samples were purified for sequencing using a QIAgen Gel Extraction Kit (QIAgen), according to the instructions of the supplier and sequenced using an Applied Biosystems 3730xl sequencer at the University of Minnesota Genomics Center. To confirm that a sequenced gene had orthology to a published *lsr2*, the sequences were entered into the BLASTn server. If a sequence showed orthology to a published *lsr2* gene with a 70% or better match it was verified by multiple species alignment against several other published or sequence-identified *lsr2* genes.

Phylogenetic analysis

Before phylogenetic analysis was attempted, the *lsr2* fragments were trimmed to exclude the PCR primer binding sites. Sequences were trimmed before the 5' end consensus sequence GAATTC (position 69 in the *M. smegmatis lsr2* gene) and after the 3' end consensus sequence GTGTCGA (position 323 in the *M. smegmatis lsr2* gene) using the Jalview program. The resulting fragments were roughly 250 bp in length and preserved the original reading frame when translated. Phylogenetic trees were constructed with the MEGA program using the neighbor-joining method. The stability of the trees was evaluated using boostrap resampling with 1,000 replicates.

RESULTS

Analysis of PCR products

PCR of the *lsr2* gene resulted in amplification of a 325-bp product in all 54 *Mycobacterium* strains tested. No amplification of '*lsr2'* was observed in control strains of *E. coli.* (data not shown). Figure 2 indicates that DNA synthesis errors made by *taq* polymerase were not a likely source of error in this experiment, as three separate amplifications of *lsr2* were identical.

Sequencing and subsequent analysis of the PCR products revealed that 48 of the samples showed some identity to *lsr2* genes that were currently published on the NCBI database. Sequence similarity to published *lsr2* ranged from 79%-100%, drawing matches for *M. tuberculosis, M. smegmatis, M. gordonae* and others. These PCR sequences were aligned to highlight nucleotide differences in Figure 3. Translation of the PCR fragments showed that there was variation at the amino acid level in the (Figure 4), with 19 residues that showed a degree of variability between species.

Phylogenetic analysis

A phylogenetic tree was constructed using 48 *Mycobacterium lsr2* genes from PCR sequencing as well as published sequences from the NCBI database (Figure 1). *E. coli h-ns* gene sequences were used as an outgroup, due to known structural and functional homologies with Lsr2 (Gordon, 2008), and the sequences were constructed into a tree using the neighbor joining method. The resulting phylogentic trees showed no differences between several members of the *M. tuberculosis complex* (*M. tuberculosis,*

M. canettii, *M. africanum*), no difference between *M. marinum* and *M. ulcerans*, no difference between *M. senegalense* and *M. peregranum*, and no difference between *M. goodii* and *M. cookii*. A pairwise alignment (Table 2) of *lsr2* sequence differences shows that on average there is a 21.9% difference between species, with a maximum difference of 45.5% between *M. chitae* and *M. abscessus*. Bootsrap values were included to show the percentage of instances that certain species were clustered together out of 1000 replicate trees generated by the program.

DISCUSSION

Despite the finding that Lsr2 regulates biofilm formation in *M. smegmatis*, little has been done to investigate the prevalence and role of Lsr2 in other *Mycobacterium*. It is known that *lsr2* is an essential gene in *M. tuberculosis*, however its role is still unclear (Sassetti, 2003). At the onset of this work, there were only 12 identified *lsr2* sequences on the NCBI database, mostly as a result of whole-genome sequencing. The purpose of this work was to amplify *lsr2* from 52 *Mycobacterium* species using PCR in order to assess the prevalence and diversity of the gene in *Mycobacterium*.

The *lsr2* gene was positively identified in 46 of the 52 species screened. The 6 species that produced no '*lsr2'* PCR products were *M. dienhoferi, M. alvei, M. palustre, M. chubuense, M. trivale,* and *M. mucogenicum* (Figure 5). It is possible that these species do not have Lsr2, however it is also possible that the PCR primers were unable to amplify *lsr2* in those species due to differences in the primer binding regions. Southern blotting might indicate that the gene is present in these species. However if the PCR primer

binding regions have sufficient differences to prevent amplification then it is also possible that there is sequence dissimilarity in these species to the point that DNA-DNA probing would also be difficult. An alternative approach might be to introduce a functional copy of *lsr2* to these 6 species via multicopy plasmid or integrative vector. Perhaps changes in biofilm formation or colony morphology phenotypes would be observed, indicating that Lsr2 was not native or functional and the introduced protein has influenced the expression of genes.

The *lsr2* gene is a relatively small gene in *M. smegmatis* (345 bp long), however, this is a size that appears to be slightly variable between species with common published gene sizes of between 342 bp - 339 bp. A region within the 325-bp PCR fragment generated in this study was analyzed for its utility as a way to identify unknown mycobacterial isolates. This region ranged from 218-224 bp long and was chosen because it was internal to the PCR primer binding sites. Additionally, this region was far enough downstream of the sequencing primer site to allow for stabilization of the sequencing reaction when sequencing from a PCR product. This allowed for sequence identification of the species without the need to first ligate the PCR fragment into a plasmid prior to sequencing. Phylogenetic analysis of this fragment was capable of distinguishing between all *Mycobacterium* strains tested with the exceptions of no discernible differences between *M. tuberculosis, M. canettii,* and *M. africanum,* between *M. marinum* and *M. ulcerans,* between *M. senegalense* and *M. peregranum,*and between *M. goodii* and *M. cookii.* All other species are observed to have interspecies differences ranging from 0.5%-45.5% with an average difference of 21.9% between species. These

differences are shown in Table 2, which displays a pairwise comparison between all 52 species tested. These results indicated that *lsr2* shows a high degree of interspecies variability on par with current sequencable markers used to identify *Mycobacteria* to the species level.

Traditionally the 16S rRNA gene has been used to identify bacterial species based on sequence. However, this gene does not appear to be polymorphic enough to confidently identify *Mycobacterium* to the species level (Yamada-Noda, 2007). To address this problem several genes such as *dnaJ, hsp65, rpoB* and *secA* have all been identified as polymorphic sequencable markers specific to *Mycobacterium* (Yamada-noda, 2007)(Ringuet, 1999)(Zelazny, 2005)(Kim, 1999). The *dnaJ* gene is commonly used for differentiating species of *Mycobacterium* because the 365-bp PCR fragment of *dnaJ* has an average difference of 24.3% between species. The '*lsr2'* gene sequence shows comparable differences between *Mycoabcterium* species with an average difference of 21.9%. If it were possible to analyze the entire 350-bp *lsr2* gene it may yield higher resolution, however, it is difficult to design universal PCR primers for this task and complete *lsr2* sequences are not available on GenBank for most *Mycobacterium* species. The partial sequences generated in this work represent the only survey of *lsr2* among the less common members of the *Mycobacterium* genus.

Interestingly, there appears to be an EcoRI restriction site (GAATTC) located in nearly every *lsr2* sequence that was studied. This site occurs from position 1-6 in the multiple species alignment in Figure 3. This was considered odd since *lsr2* is particularly GC rich

(67% GC) and yet the EcoRI site was among the largest and most conserved regions in the gene alignments. When the *lsr2* PCR product was digested with EcoRI the fragment was clearly cut into two pieces, one at ~250 bp and the other ~70 bp (data not shown). This served as an excellent test to determine if the correct PCR fragment had been isolated and was useful for deciding if the fragment was a candidate for sequencing.

Nucleotide differences do not necessarily manifest as phenotypic differences at the primary structure levels of proteins, so differences in *lsr2* may not necessarily change the primary structure of the Lsr2 protein. To evaluate this observation the nucleotide sequences for '*lsr2*'were translated into amino acid residues in Figure 4 using MEGA. When Figure 3 and Figure 4 were cross-referenced it became apparent that most of the variability that was seen in the nucleotide sequences was located at the third nucleotide of each codon, resulting in mostly silent mutations at the amino acid level. There were 19 residues that showed 10% or greater variability between species in the amino acid alignment, including a region from positions 44-46 that seemed to be absent to varying degrees in many of the species (Figure 4). This information is interesting, but inconclusive as to whether there are functional differences amongst the various species of *lsr2.* Because the regions studied were internal fragments it was not possible to superimpose the variable regions onto a 3-dimensional model that represented a complete protein. It would, however, be possible to align these partial sequences with a known 3D model of Lsr2 to see if there is a specific area of the folded protein that harbors an abundance of these variations. A future direction for this research would be to sequence the entire *lsr2* gene for each species and use the data to determine the location of these

possible hypervariable sites in the folded protein structure.

The highly conserved nature of *lsr2* among the 54 *Mycobacterium* tested indicates that this is an evolutionarily old gene. *Mycobacterium* genomes can vary widely between species, and yet the *lsr2* fragment was found in 48 species with little change in size or translated amino acid structure (Fig. 3 and Fig. 4). The observation that Lsr2 primary structure varies little between species suggests a vital role in the genus. A 2003 study by Sassetti found lsr2 to be an essential gene is M. tuberculosis, although this is not this case for *M. smegmatis* (Chen, 2006). Future studies could focus on determining the role that Lsr2 plays in a wider variety of *Mycobacteria*, and perhaps further investigate the ways in which Lsr2 interacts with the genome.

REFERENCES

Arora, K., D. C. Whiteford, D. Lau-Bonilla, C. M. Davitt, and J. L. Dahl. 2008. Inactivation of *lsr2* results in a hypermotile phenotype in *Mycobacterium smegmatis*. J. Bacteriol. 190:4291–4300.

Chen, J., German, G., Alexander, D., Ren, H., Tan, T., and Liu, J. 2005. Roles of Lsr2 in Colony Morphology and Biofilm Formation of Mycobaterium smegmatis. Journal of Bacteriology. 633-641.

Chen, J. M., H. Ren, J. E. Shaw, Y. J. Wang, M. Li, A. S. Leung, V. Tran, N. M. Berbenetz, D. Kocincova, C. M. Yip, J. M. Reyrat, and J. Liu. 2008. Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. Nucleic Acids Res. 36:2123–2135.

Colangeli, R. ,A. Haq, V. L. Arcus, E. Summers, R. S. Magliozzo, A. McBride, A. K. Mitra, M. Radjainia, A. Khajo, W. R. Jacobs, Jr., P. Salgame, and D. Alland. 2009. The multifunctional histone-like protein Lsr2 protects *mycobacteria* against reactive oxygen intermediates. Proc. Natl. Acad. Sci. U. S. A. 106:4414–4418.

Colangeli, R., D. Helb, C. Vilcheze, M. H. Hazbon, C. G. Lee, H. Safi, B. Sayers, I. Sardone, M. B. Jones, R. D. Fleischmann, S. N. Peterson, W. R. Jacobs, Jr., and D. Alland. 2007. Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in *M. tuberculosis*.

Falkinham JO 3rd, Norton CD, LeCheva- llier MW. 2001. Factors influencing numbers of *Mycobacterium avium, Mycobacterium intracellulare*, and other *Mycobacteria* in drinking water distribution systems. Appl. Environ. Microbiol. 67:1225–31

Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. 2009. Opportunistic pathogens enriched in showerhead biofilms. Proceedings of the National Academy of Sciences U S A.;106:16393–9

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

Gordon BR, Imperial R, Wang L, Navarre WW, Liu J. 2008. Lsr2 of Mycobacterium represents a novel class of H-NS-like proteins. Journal of bacteriology 190: 7052–7059

Gordon, B. R., Y. Li, L. Wang, A. Sintsova, H. van Bakel, S. Tian, W. W. Navarre, B. Xia, and J. Liu. 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. 107:5154–5159.

Hall-Stoodley, L., and H. Lappin-Scott. 1998. Biofilm formation by the rapidly growing *mycobacterial* species Mycobacterium fortuitum. FEMS Microbiol. Lett. 168:77–84.

Hall-Stoodley, L., J. W. Costerton, and P. Stoodley. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat. Rev. Microbiol. 2:95–108.

Hall-Stoodley, L., Brun, O.S., Polshyna, G., and Barker, L.P. 2006 *Mycobacterium marinum* biofilm formation reveals cording morphology. FEMS Microbiol Lett 257: 43–49.

Kim B. J., Lee S. H., Lyu M. A., Kim S. J., Bai G. H., Chae G. T., Kim E. C., Cha C. Y., Kook Y. H. 1999 Identification of *mycobacterial* species by comparative sequence analysis of the RNA polymerase gene (rpoB). J. Clin. Microbiol. 37:1714–1720.

Laal, S., Y. D. Sharma, H. K. Prasad, A. Murtaza, S. Singh, S. Tangri, R. S. Misra, and I. Nath. 1991. Recombinant fusion protein identified by lepromatous sera mimics native *Mycobacterium leprae* in T-cell responses across the leprosy spectrum. Proc. Natl. Acad. Sci. U. S. A. 88:1054–1058.

Nguyen K. T., Piastro K., Gray T. A., Derbyshire K. M.. 2010. Mycobacterial biofilms facilitate horizontal DNA transfer between strains of Mycobacterium smegmatis. J. Bacteriol. 192:5134–5142.

Ojha, A. K., A. D. Baughn, D. Sambandan, T. Hsu, X. Trivelli, Y. Guerardel, A. Alahari, L. Kremer, W. R. Jacobs, Jr., and G. F. Hatfull. 2008. Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol. Microbiol. 69:164-174

Ringuet, H., C. Akoua-Koffi, S. Honore, A. Varnerot, V. Vincent, P. Berche, J. L. Gaillard, and C. Pierre-Audigier. 1999. hsp65 sequencing for identification of rapidly growing mycobacteria. J. Clin. Microbiol. 37:852-857

Sassetti, C. M. & Rubin, E. J. 2003. Genetic requirements for *mycobacterial* survival during infection. Proc. Natl Acad. Sci. USA 100, 12989–12994

Saitou N. and Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.

Stoodley, P., K. Sauer, D. G. Davies, and J. W. Costerton. 2002. Biofilms as complex differentiated communities. Annu. Rev. Microbiol. 56:187– 209.

Tamura K., Nei M., and Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (*USA*) **101**:11030-11035.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

Yamada-Noda, M., Ohkusu, K., Hata, H., Shah, M. M., Nhung, P. H., Sun, X. S., Hayashi, M. & Ezaki, T. 2007. Mycobacterium species identification – a new approach via dnaJ gene sequencing. Syst Appl Microbiol 30, 453–462.

Zelazny, A. M., Calhoun, L. B., Li, L., Shea, Y. R. & Fischer, S. H. 2005.

Identification of Mycobacterium species by secA1 sequences. J Clin Microbiol 43, 1051–1058

The Δ*lsr2* phenotype of *Mycobacterium smegmatis* is linked to genes associated with mycocerosic acid synthase

ABSTRACT

Mycobacterium smegmatis is a non-pathogenic model organism commonly used to study the human pathogen *Mycobacterium tuberculosis.* *M. smegmatis* has a rough colony morphology and is capable of forming biofilms, however, when the DNA-regulating protein Lsr2 is inactivated the bacteria forms a smooth colony and loses biofilm formation. This study investigates the phenotypic effect on colony morphology and biofilm formation when Lsr2-regulated genes are inactivated. A mutant strain of *M. smegmatis* with defective *lsr2 (M. smegmatis* Δ*lsr2)* was used as a parent strain for the study and the gene for mycocerosic acid synthase (*mas*) was selected for inactivation. Genes downstream of *mas* from MSMEG_4727 to MSMEG_4733 and MSMEG_4741were also selected for inactivation. Genes were inactivated using homologous recombination via "suicide vector" plasmids lacking a *Mycobacterial* origin of replication and carrying kanamycin resistance and PCR fragments of the target genes. *M. smegmatis* Δ*lsr2* was transformed by electroporation and transformants were plated and screened for kanamycin resistance. The mutant strain lacking functional *lsr2* and *mas (M. smegmatis* Δ*lsr2*Δ*mas)* regained the ability to form biofilms and exhibited colony morphology similar to wild-type *M. smegmatis.* Inactivation of genes MSMEG_4727 to MSMEG_4733 produced similar results. Inactivation of MSMEG_4741 caused no change in the Δ*lsr2* mutant phenotype. These results indicate

that *mas* and genes downstream are involved with the biofilm formation and colony morphology change associated with the *M. smegmatis* Δ*lsr2* mutant strain.

INTRODUCTION

Many species of *Mycobacterium* are known to form biofilms, which are a matrix of extracellular polymeric substances (EPS) (Wingender, 1999) excreted by the cells to aid in attachment to surfaces (Branda et al., 2005)(Hall-Stoodley, 2004) (Hall-Stoodley, 2002). Biofilm forming bacteria are less susceptible to antibiotics (Colangeli, 2007), (Stewart and Costerton, 2001) and can colonize surfaces allowing bacteria to persist in flowing liquid systems (Geesey, 1977).

Mycobacterium smegmatis is commonly used to study biofilm formation among the members of the genus *Mycobacterium*. *M. Smegmatis* is a popular model system for studying *mycobacterial* phenomena because of its short doubling time, ease of culture, and genetic similarity to *Mycobacterium tuberculosis.* When grown in liquid media lacking a detergent, *M. smegmatis* forms a biofilm along the sides of the culture vessel and forms a floating raft of cells called a pellicle at the air-liquid interface (ALI). Additionally, many of the hydrophobic, planktonic cells will clump together and sink to the bottom of the liquid. This hydorphobicity is one of the traits of mycolic acids and surface molecules that define the *Mycobacterium*.

Biofilm formation in bacterial strains such as *Escherichia coli, Pseudomonas aeruginosa,* and *Vibrio cholera* has been found to be linked to certain genes, (Heilmann, 1996)(Mack,

1994)(O'Toole and Kolter, 1998) However, little is known about what genes regulate *mycobacterial* biofilms. One recently studied biofilm-dependent gene is *lsr2,* which codes for the DNA bridging protein Lsr2. The Lsr2 protein was originally described in *Mycobacterium leprae,* where it was identified as an immunological antigen (Laal et al 1991). Since then, Lsr2 has been characterized in many other *Mycobacterium* (Gatlin, unpublished results), and its role in *M. smegmatis* has been studied in great detail (Arora, 2008)(Chen, 2005)(Chen, 2006)(Colangeli, 2009).

In 2006, Chen *et al.* reported that transposon mutagenesis of *lsr2* caused *M. smegmatis* to lose its ability to form biofilms or pellicles. This mutant strain of Lsr2-deficient bacteria was referred to as *M. smegmatis* Δ*lsr2,* because the *lsr2* gene was inactivated. Further studies show that *M. smegmatis* Δ*lsr2* has an increased ability to spread over soft agar surfaces; a trait that was recently discovered in wild-type (WT) *M. smegmatis* (Chen, 2006)(Arora, 2008). Colony morphology is also affected by *lsr2* inactivation, *lsr2* mutants displaying a smooth shiny texture in contrast to the rough matte colonies of the isogenic WT strain. The morphologies of WT and Δ*lsr2* colonies are shown in Figure 6.

More recently it was found that Lsr2 had the ability to act as a transcriptional regulator in *M. smegmatis*, and inactivation of *lsr2* can increase or decrease expression of over 60 genes (Colangeli, 2007). One gene regulated by Lsr2 is called *mycocerosic acid synthase* (*mas*), which encodes a multifunctional enzyme that catalyzes the synthesis of very long chain multiple methyl branched fatty acids called that are part of the biosynthetic pathway of lipoologisaccharides in the cell wall (Etienne, 2007). These fatty acids are

expected to be on the surface of *M. smegmatis* cells and may be directly linked with cell hydrophobicity and biofilm formation (Lemassu, 1992). In the *M. smegmatis* Δ*lsr2* strain, *mas* is expressed at a level 5-fold higher than in the WT strain (Colangeli, 2007). This suggests that in WT cells, Lsr2 functions as a repressor of *mas* expression. We hypothesize that inactivation of Lsr2 leads to elevated expression of *mas,* which causes smooth colonies and cells that cannot form biofilms.

This research tests the hypothesis that *mas* gene expression contributes to change in phenotype observed in *M. smegmatis* Δ*lsr2*. In the studies described below, *mas* and genes downstream were inactivated in *M. smegmatis* Δ*lsr2* in an attempt to eliminate the effect of increased expression of *mas*. Our results show that elevated mas expression may be involved in colony morphology and loss of biofilm formation seen in *M. smegmatis* Δ*lsr2*, but this gene is likely not the only one controlling the phenotype of *M. smegmatis* Δ*lsr2*.

MATERIALS AND METHODS

Bacterial strains

A list of bacterial strains and plasmids used in this study is shown in Table 3. Normal cultivation of *M. smegmatis* was carried out at 37˚C on Middlebrook 7H11 (Difco) agar medium, and liquid cultures were grown in 7H9 (Difco) supplemented with 0.05% Tween 80. Bacterial strains were transformed by electroporation as previously described

(Wei, 2000). Transformants were selected on 7H11 (Difco) agar plates supplemented with 10% OADC, 0.3% glycerol, and kanamycin (25µg/mL).

Generation of suicide vectors

Genetic knockout variants of *M. smegmatis* were generated by introduction of a suicide vector carrying a PCR-generated internal fragment of the gene to be deactivated. This plasmid vector lacks an origin of replication in *Mycobacterium* and so the potential for kanamycin-resistance in *mycobacterial* transformants necessitates integration ("suicide") of the plasmid into the chromosome. This integration is most likely to occur in homologous regions shared by chromosome and plasmid. To inactivate *mycobacterial* genes, the first step was PCR amplification of an internal fragment of the gene to be deactivated. Primers were designed to amplify a region of the gene that was missing significant portions of the 5'and 3' ends. A complete list of PCR primers and products is shown in Table 2. Truncated gene fragments were ligated into pTOPO (Invitrogen) and transformed into OneShot TOP10 chemically competent *E. coli* (Invitrogen) according to supplier recommendations. Successful ligations were selected by blue/white screening and digested with BstXI to confirm correct product insertion.

PCR

All PCR was performed using Illustra PuReTaQ ready-to-go PCR beads in 0.2mL tubes. A complete list of PCR primers and templates is shown in Figure 2. PCR cycles used are the following:

Internal *mas* cluster fragment PCR:

95˚C for 5 min, 35 cycles of 97˚C for 30 sec, 62˚C for 30 sec, 72˚C for 1 min and a final 72˚C for 5 min.

mas PCR

95˚C for 5 min, 35 cycles of 97˚C for 30 sec, 62˚C for 30 sec, 72˚C for 4 min and a final 72˚C for 5 min.

lsr2 PCR

95˚C for 5 min, 35 cycles of 95˚C for 15 sec, 59˚C for 30 sec, 72˚C for 1 min and a final 72˚C for 5 min.

Subcloning PCR Fragments

PCR fragments were gel purified using QIAgen preps and verified to be the correct size by gel electrophoresis. Fragments were subcloned into pTOPO according to the supplier's directions, and the resulting recombinant plasmids used to transform chemically competent *E. coli* that were allowed to grow on LB plates with kanamycin and X-gal. Plasmids were isolated from white colonies of *E. coli* , and were verified to contain proper inserts by restriction analysis and gel electrophoresis.

Creating electrocompetent *M. smegmatis*.

M. smegmatis was made electrocompetent as previously described (Wei, 2000). Briefly, cells were grown to early log phase (OD_{500} of 0.5) while shaking in 7H9 supplemented with 0.05% tween 80 at 37˚C. Cells at this density were immediately placed on ice for 10 minutes and aliquoted into 50-mL Falcon tubes. Tubes were centrifuged at 2,500 xg for

10 minutes at 4˚C to pellet cells, and the supernatant was removed. Cells were resuspended in 50 mL of ice-cold 10% glycerol solution and chilled in an ice bath without agitation for 15 min. A second wash was done by re-pelleting the cells as before, removing the supernatant, replacing with 50 mL of fresh ice-cold glycerol solution, and incubating without agitation for 15 min on ice. For the final wash the cells were pelleted and resuspended in 1-mL glycerol solution was added. Cells were used immediately for electroporation or stored frozen at -80˚C.

Transformation of *M. smegmatis*

Electrocompetent *M. smegmatis* was transformed by electroporation with a BioRad electroporator. 300-µL aliquots of electrocompetent cells were loaded into Biorad Gene Pulser 0.2-cm gap cuvettes along with 4 µL of plasmid construct. Samples were electroporated at 2.5 KV, 1000 Ω (resistance), 25µF (capacitance) and immediately suspended in 1 mL of 7H9 supplemented with 0.05% Tween 80. Cells were allowed to recover for 4 h at 37˚C shaking at 200 RPM and plated on 7H11 agar supplemented with kanamycin (50 µg/mL). Plates were incubated for 72 hours at 37˚C before screening for transformants. Isolated colonies were selected and streaked on fresh 7H11 supplemented with kanamycin (50 µg/mL).

Verification of homologous recombination

Homologous recombination as a result of transformation was verified by PCR amplification of affected regions of DNA using primer pairs flanking the insertion site. A forward PCR primer was designed to bind a chromosomal region upstream of the

truncated gene fragment and a reverse primer was designed to bind downstream of the truncated fragment. PCR amplification was done with using these primer pairs on the transformation candidates with the expectation that insertion of the pTOPO backbone would increase the distance between the primer pairs by 4000 bp relative to a PCR product of the native gene. Transformant PCR products were compared side-by-side with the PCR product from the native gene by gel electrophoresis. The "*mas* PCR' cycle described above was used with each set of primers.

Colony Morphology analysis

M. smegmatis Motility assays were performed as previously described (Arora, 2008). Briefly, cells were cultured in shaking 7H9 liquid medium at 37˚C until late log phase and a 2-µL aliquot was spotted onto medium consisting of M63 minimal salts media (Difco) supplemented with 0.5% casamino acids (w/v), 0.2% glycerol, 1mM $MgCl_2$ and solidified with agarose at a final concentration of 0.3%, 0.8%, or 1.5% (w/v). Inoculated plates were kept humidified at 37˚C in a plastic bag with a moist paper towel for 48 h and then photographed.

Cell clumping, pellicle formation, and biofilm assays

The ability of *M. smegmatis* strains to clump and form biofilms and pellicles at the air-liquid interface (ALI) was analyzed in 7H9 liquid media, as previously described (Arora, 2008). Briefly, cells were first grown to mid log phase in 7H9 with or without 0.05% Tween 80 at 37˚C. These cultures were used to inoculate culture tubes containing 3 mL

of 7H9 with or without 0.05% Tween 80. The culture tubes were left undisturbed for 48 h at 37˚C and then photographed.

Complementation of mas

PCR was used to amplify an 8.6-Kb region of DNA that contained mas and a 1000-bp non-coding region that preceded it. This DNA fragment was ligated into the expression vector pSE100, which contains an origin of replication for *Mycobacterium*, Hygromycoin resistance gene, and a tetracycline-induced promoter region. The 8.6-Kb *mas* was ligated in downstream of the promoter region using an In-Fusion Cloning kit from Clontech and following the manufacturers instructions. *M. smegmatis* Δ*lsr2*Δ*mas* was made electrocompetent and transformed with pSE100::mas using electroporation as described earlier in this work. Transformants were selected on 7H11 agar plates supplemented with kanamycin (50 µg/mL) and hygromycin (25 µg/mL) and allowed to grow for 72 hours at 37˚C after plating. Colonies were re-streaked and screened for the plasmid by using PCR primers specific to the plasmid backbone. Verified *M. smegmatis* Δ*lsr2*Δ*mas* pSE100::*mas* transformatns were plated on 7H11 agar plates supplemented with kanamycin (50 µg/mL) and hygromycin (25 µg/mL) and tetracycline (100µg/mL) and grown for 72 hrs. Additional plates supplemented with tetracycline concentrations of 10 µg/mL, 50 µg/mL, 200 µg/mL, 500 µg/mL 750 µg/mL, 1000 µg/mL, and 5000 µg/mL were used. Colonies were analyzed for morphology and biofilm formation as described previously in this work.

RESULTS

The basis of this research was to understand the molecular reasons for a phenotypic change in biofilms that occurs in *M. smegmatis* when the *lsr2* gene is inactivated. Wild-type (WT) *M. smegmatis* forms rough, opaque colonies on 7H11 agar and will form a biofilm with a floating pellicle in 7H9 liquid media. When the *lsr2* gene is inactivated (Δ*lsr2*), mutant *M. smegmatis* colonies take on a smooth, shiny texture and the community of cells loses the ability to form a biofilm and pellicle in liquid media.

Previous expression studies showed that when *lsr2* was inactivated genetic regulation of about 60 genes changed, in some cases with increased expression up to 5-fold in the absence of *lsr2* (Colangeli, 2007). We hypothesized that genes normally repressed by Lsr2 are involved in the smooth colony morphology and lack of biofilm formation seen in the *M. smegmatis* Δ*lsr2* strain. This study explores the effect on colony morphology and biofilm formation when individual Lsr2-repressed genes in the Δ*lsr2* mutant of *M. smegmatis* are inactivated.

The first Lsr2-repressed gene selected for study was MSMEG_4727, which has the putative function of mycocerosic acid synthase (*mas*). The hypothesis was that inactivation of *mas* in the Δ*lsr2* strain of *M. smegmatis* would lead to a phenotype where the bacteria regained the ability to form biofilms and the rough colony morphology. To test the hypothesis, *M. smegmatis* with a truncated, nonfunctional *lsr2* gene was used as a parent strain for inactivation of *mas*. This parent strain is referred to as *M smegmatis*

Δ*lsr2* and it is unmarked, meaning that it does not contain hygromycin or kanamycin resistance genes.

To selectively inactivate *mas*, an internal fragment of the native *mas* gene, which was truncated at the 5' and 3' ends, was amplified by PCR and cloned into pTOPO. The resulting suicide vector was named pTOPO::'*mas*', and had the following characteristics: it lacked an origin of replication in *Mycobacterium* (OriM) and contained the *aph* gene for kanamycin resistance. A cartoon schematic of this construct and its relation to *mas* can be seen in Figure 7A.

The pTOPO::'*mas*' construct was used to transform *M. smegmatis* Δ*lsr2* and colonies were selected for on 7H11 plates containing kanamycin. Because the construct lacked OriM, pTOPO::*'mas'* was not able to replicate autonomously in the *Mycobacterium* cell, and would not be passed to daughter cells in a replicating colony. The only condition that would allow for a colony to form on the kanamycin plates was integration of pTOPO::'*mas*' into the *mycobacterial* chromosome via homologous recombination between '*mas'* and the native *mas* gene. This recombination resulted in an inactive meridiploid version of *mas* consisting of two partial *mas* genes separated by the backbone of pTOPO (Fig. 7B). One truncated *mas* gene lacks a start codon, and the other is significantly shortened at the 3' end. In this meridiploid condition two inactive forms of *mas* exist on the chromosome.

Inactivation of *mas* was verified by PCR analysis as shown in Figure 7C. Primer pairs were used to show that an insertion had taken place by pairing a primer that binds to the backbone of the pTOPO plasmid (M13F or M13R) with a primer that binds to *mas* (masF or masR). A transformation candidate colony's genomic DNA serves as the template. If recombination took place, the M13 primer would be able to bind and in conjunction with the *mas* primer create a PCR fragment that represented one of the partial mas fragments created during homologous recombination. This was done for both partial gene fragment's which can be seen in lanes 3 and 4 of the gel image in Figure 7C. When these primer combinations are used on WT genomic DNA, there is no binding site for the M13 primer and so no amplification is observed in lanes 7 and 8 of Figure 7C.

Additionally, primers designed to amplify the entire 7.6 Kb *mas* gene were used to assay for transformation. With WT DNA as template, masF and masR can amplify a 7.6-Kb fragment, which is seen in lane 2 of Figure 7C. When the pTOPO plasmid is inserted mid-gene into *mas* during homologous recombination, the distance between masF and masR increases to 11.4 Kb, which is beyond the capabilities of the PCR amplification conditions used in this study. Consequently, no band is observed in lane 6 of Figure 7C.

Inactivation of *mas* from *M smegmatis* strain Δ*lsr2* resulted in a strain that was referred to as *M. smegmatis* strain Δ*lsr2*Δ*mas*, since it contained nonfunctional copies of both genes. The colony morphology changed from the smooth texture of the Δ*lsr2* strain to a rough texture that resembled the WT (Fig. 9). Additionally, the Δ*lsr2*Δ*mas* had the ability to form biofilms, a characteristic of WT.

Although the results of Figure 9 seemed to indicate that *mas* has an effect on biofilm formation and colony morphology, verification would require complementing the Δ*lsr2*Δ*mas* mutant with an extrachromosomal copy of *mas* on an expression vector. No changes from the Δ*lsr2*Δ*mas* mutant phenotype were observed for colony morphology or biofilm formation during complementation attempts (data not shown). Because the plasmid containing full-length *mas* could not complement the loss of *mas* in *M. smegmatis* Δ*lsr2*Δ*mas*, we hypothesized that the *M. smegmatis* Δ*lsr2*Δ*mas* phenotype was due to a polar effect of inserting plasmid DNA into the *mas* gene locus. In disrupting *mas* with the suicide vector, we may have inadvertently disrupted expression of genes downstream of *mas* that are the real explanation of why the *M. smegmatis* Δ*lsr2*Δ*mas* strain has a WT colony morphology and normal biofilm formation. Therefore, an approach was taken to investigate the possibility of a polar effect. To challenge the hypothesis that *mas* is responsible for the Δ*lsr2* phenotype, a genetic region 1000 bp downstream of *mas* was inactivated from the Δ*lsr2* strain using the same technique as was used to inactivate *mas* (Fig 8). This region is called MSMEG_4728, and has an identified start and stop codon set, but no putative function. Inactivating MSMEG_4728 had the same effect on colony morphology and biofilm formation as inactivating *mas* (Fig 9*)*; The transformed bacteria closely resembled the Δ*lsr2*Δ*mas* phenotype.

Putative genes shown in Figure 10 were inactivated individually in the same manner as *mas* and MSMEG_4728. Beginning with MSMEG_4729, each of the genes was individually inactivated sequentially in the parent Δ*lsr2* strain by insertion of pTOPO

suicide vectors into these open reading frames. One prediction was that the endpoint of the *mas* cluster would be MSMEG_4733, because this gene was the first of many oriented in the opposite transcriptional direction as *mas*. This led to an experimental design that targeted MSMEG_4729-4733 for sequential, individual inactivation.

The phenotypic results of inactivating MSMEG_4728 can be seen in Figure 9. Similar phenotypes were observed in the Δlsr2Δ4727-Δ4733 strains (data not shown), which indicates there is some common effect on the colony morphology and biofilm formation among these genes. Inactivation of Δ4741 has no effect on the *M. smegmatis* Δ*lsr2* phenotype (data not shown). This indicates that MSMEG_4741 is a gene that lies outside of a "mas cluster" that regulates colony morphology and biofilm formation in *M. smegmatis*

It was necessary to make sure that the revision to WT seen in this study was not an effect of introducing the pTOPO backbone into the *M. smegmatis* chromosome. We wished to perform a control experiment to test if the insertion of pTOPO suicide vectors into the chromosome of *M. smegmatis* Δ*lsr2* had some effect on colony morphology and biofilm formation. Specifically, we wanted to make sure that the phenotype change from Δlsr2 back to the WT-like phenotype wasn't some effect of introducing the pTOPO backbone into the *M. smegmatis* chromosome. The hypothesis was that a pTOPO backbone integrated into the chromosome in a location neutral to genes of the *mas* cluster would not have an effect on the Δ*lsr2* phenotype. The gene MSMEG_6092 (*lsr2*) was selected as a neutral site for pTOPO insertion in this experiment because MSMEG_6092 was

already inactive in the Δ*lsr2* parent strain. Further manipulation of an already inactive gene should not produce any effects due to changed gene expression, allowing the pTOPO insertion to be the only variable. The hypothesis of the control experiment was supported; introduction of pTOPO into the inactive *lsr2* did not have an effect on the Δ*lsr2* phenotype. Colony morphology of the resulting transformants resembled the Δ*lsr2* parent strain in regards to colony morphology and biofilm formation.

DISCUSSION

This work describes a cluster of genes that appears to play a role in the ability of *M. smegmatis* to form colonies and biofilms in the absence of the DNA bridging protein Lsr2. The Δ*lsr2* mutant of *M. smegmatis* has been studied previously for its changes in gene expression, colony morphology, and biofilm formation (Chen, 2005)(Arora, 2008). Wild-type *M. smegmatis* normally forms dry, flat colonies with irregular, lobate margins and liquid cultures of the bacteria are able to form a biofilm on the side of a test tube. When *lsr2* is inactivated the bacteria form a shiny, rounded colony with smooth margins and the bacteria are not able to form a biofilm in liquid media. This study aimed to identify specific genes that were involved in the phenotype change in the absence of Lsr2.

Previous studies indicated that Lsr2 is a DNA bridging protein that has the ability to act as a transcriptional activator or repressor (Gordon, 2010), and have identified regions of the genome that are affected (Colangeli, 2007). The study by Colangeli in 2007 indicated that Lsr2 repressed MSMEG_4727-MSMEG_4736 in WT M. smegmatis, where

MSMEG_4727 was the gene for mycocerosic acid synthase (*mas*). The putative function of *mas* is polyketide synthesis, and studies of *mas* in WT *M. smegmatis* have shown that the bacteria lose the ability to produce lipooligosaccharides when *mas* is inactivated (Etienne, 2009). Lipooligosaccharies are antigenic compounds found in *M. marinum* and other *Mycobacterium* and are involved in the formation of biofilms, sliding motility, and macrophage infection (Ren, 2007). This previous work indicated that *mas* was a good candidate gene for inactivation in the Δlsr2 strain as a starting point for our study.

When Lsr2 is absent from *M. smegmatis*, *mas* is no longer repressed and the gene is expressed at a much higher rate. We hypothesized that this higher rate of *mas* expression created the Δ*lsr2* phenotype. The prediction was that if increased expression of *mas* in Δ*lsr2 M. smegmatis* caused the loss of biofilm formation and changes in colony morphology, then inactivating *mas* in the Δ*lsr2* mutant would restore biofilm formation and colony morphology to a state similar to WT. Inactivating overexpressed *mas* was expected to mimic the effect of normal WT Lsr2 repression of *mas*, causing the bacteria to "revert" back to a WT phenotype. The results supported this prediction, showing that inactivating *mas* does create a biofilm-forming Δ*lsr2* strain similar to the WT that we called Δ*lsr2*Δ*mas*. The Δ*lsr2*Δ*mas* strain was able to synthesize a biofilm that resembled the WT and also formed dry, flat colonies with irregular, lobate margins (Fig. 6).

Attempts to complement Δ*lsr2*Δ*mas* with an extrachromosomal copy of *mas* were not successful, so an alternative approach to investigating a possible polar effect of *mas* inactivation was taken. The genes immediately downstream of *mas* were inactivated

individually using the same technique as before and analyzed for the ability to produce biofilms and the rough colony morphology. The prediction was that if the phenotype observed from *mas* inactivation was a polar effect, then we would see the same phenotype resulting from the downstream inactivation as Δ*lsr2*Δ*mas*. A genetic region identified as MSMEG_4728 was immediately downstream of *mas* and was selected for inactivation because it represented the next putative gene in the sequence.

The results of inactivating MSMEG_4728 also produced the same phenotypic effect as inactivating *mas* (Fig. 8). The transformed Δ*lsr2*Δ*4728 M. smegmatis* was able to form biofilms and exhibited rough colony morphology resembling the WT. This finding could indicate that the Δ*lsr2*Δ*mas* phenotype was a polar effect, but could also indicate that MSMEG_4728 was equally important for the Δ*lsr2* phenotype as *mas*. The previous studies of Etienne, 2009 indicated that other genes in the *mas* cluster produced proteins that were associated with the polyketide synthase coded by *mas*, including the putative acetyltransferase coded by MSMEG_4728. They suggested that several other genes in the *mas* cluster from MSMEG_4727 to MSMEG_4740 could also play a role in assisting the polyketide synthase or transporting lipooligosaccharies to the cell membrane. This could explain the finding that the phenotype of Δlsr2Δ4728 closely resembles that of Δ*lsr2*Δ*mas*.

To further investigate this effect MSMEG_4729 to MSMEG_4733 were also inactivated with similar results. "WT reversions" similar to those exhibited by the Δ*lsr2*Δ*mas* mutant were observed each time a gene downstream of *mas* was inactivated. This is congruent

with the findings of Etienne, 2007 where MSMEG_4731 and MSMEG_4733 were also found to be involved with products of *mas*.

To make sure that the Δ*lsr2*Δ*mas* phenotype was a response to inactivating genes that were associated with *mas* and not a polar effect, a gene downstream of the cluster identified by Etienne, 2007 was inactivated as a control. This gene, MSMEG_4741, was not associated with the polyketide synthase of *mas* and did not have a putative function. Inactivation of MSMEG_4741 did not cause a change from Δ*lsr2* phenotype, which indicates that the Δ*lsr2*Δ*mas* phenotype was not a polar effect, but was a response to the inactivation of the targeted genes in this study. We did not deem it necessary to inactivate MSMEG_4734- MSMEG_4740, as there was no need for redundant evidence that changing the expression of genes related to *mas* had an effect on the Δ*lsr2* phenotype

The overall finding of this work is that the Δ*lsr2* phenotype of *M. smegmatis* appears to be closely linked to the changes in expression of genes related to *mas*. In WT *M. smegmatis mas* codes for polyketide synthase, which is responsible for assembling secondary metabolites that are not essential for survival, but whose absence could impair normal growth and development of the cell (Minnikin, 2002). When the Mas polyketide synthase is inactived from WT *M. smegmatis* the bacteria loses its ability to form lipooligosaccharies in the cell wall (Etienne, 2007). The findings of our work complement these results. Lsr2 represses *mas* in WT M. smegmatis, causing low levels of Mas polyketide synthase expression. Under these WT conditions the bacteria exhibits normal biofilm production and rough colony morphology. When Lsr2 is inactivated in

the Δlsr2 strain *mas* is expressed at a higher level, producing more Mas polyketide synthase. This Δ*lsr2* condition is associated with a loss of biofilm formation and the smooth colony morphology. In the Δ*lsr2*Δ*mas* phenotype *mas* is inactive. Presumably no Mas polyketide synthase is produced and biofilm formation and rough colony morphology returns.

The Mas polyketide synthase is known to create polymethyl-branched fatty acids, which are part of the biosynthetic pathway for lipooligosaccharides (Lemassu, 1992)(Minnikin, 2002). The relationship between biofilm formation and lipooligosaccharides is not well-studied, but perhaps there is a link between the ability of *M. smegmatis* to form biofilms and the lipooligosaccharides influenced by Mas polyketide synthase.

REFERENCES

Arora, K., D. C. Whiteford, D. Lau-Bonilla, C. M. Davitt, and J. L. Dahl. 2008. Inactivation of *lsr2* results in a hypermotile phenotype in *Mycobacterium smegmatis*. J. Bacteriol. 190:4291–4300.

Branda SS, Vik S, Friedman L, Kolter R. 2005. Biofilms: the matrix revisited. Trends Microbiol.;13:20–26.

Chen, J., German, G., Alexander, D., Ren, H., Tan, T., and Liu, J. 2005. Roles of Lsr2 in Colony Morphology and Biofilm Formation of Mycobaterium smegmatis. Journal of Bacteriology. 633-641.

Chen, J. M., H. Ren, J. E. Shaw, Y. J. Wang, M. Li, A. S. Leung, V. Tran, N. M. Berbenetz, D. Kocincova, C. M. Yip, J. M. Reyrat, and J. Liu. 2008. Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. Nucleic Acids Res. 36:2123–2135.

Colangeli, R. ,A. Haq, V. L. Arcus, E. Summers, R. S. Magliozzo, A. McBride, A. K. Mitra, M. Radjainia, A. Khajo, W. R. Jacobs, Jr., P. Salgame, and D. Alland. 2009. The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. Proc. Natl. Acad. Sci. U. S. A. 106:4414–4418.

Colangeli, R., D. Helb, C. Vilcheze, M. H. Hazbon, C. G. Lee, H. Safi, B. Sayers, I. Sardone, M. B. Jones, R. D. Fleischmann, S. N. Peterson, W. R. Jacobs, Jr., and D. Alland. 2007. Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in *M. tuberculosis*.

Etienne G., Malaga W., Laval F., Lemassu A., Guilhot C., Daffé M. 2009. Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. J. Bacteriol. 191, 2613– 2621

Geesey GG, Richardson WT, Yeomans HG, Irvin RT, Costerton JW. 1977. Microscopic examination of natural sessile bacterial populations from an alpine stream. *Can. J. Microbiol.* 23(12):1733–36

Gordon, B. R., Y. Li, L. Wang, A. Sintsova, H. van Bakel, S. Tian, W. W. Navarre, B. Xia, and J. Liu. 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. 107:5154-5159.

Hall-Stoodley, L., and P. Stoodley. 2002. Developmental regulation of mi- crobial biofilms. Curr. Opin. Biotechnol. 13:228–233.

Hall-Stoodley, L., J. W. Costerton, and P. Stoodley. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat. Rev. Microbiol. 2:95–108.

Heilmann C, Gerke C, Perdreau-Remington F, Gotz F. 1996. Characterization of Tn917 insertion mutants of *Staphylococcus epidermidis* affected in biofilm formation. *Infect. Immun.* 64(1):277–82

Laal, S., Y. D. Sharma, H. K. Prasad, A. Murtaza, S. Singh, S. Tangri, R. S. Misra, and I. Nath. 1991. Recombinant fusion protein identified by lepromatous sera mimics native *Mycobacterium leprae* in T-cell responses across the leprosy spectrum. Proc. Natl. Acad. Sci. USA 88:1054–1058.

Lemassu, A., V. V. Levy-Frebault, M.-A. Laneelle, and M. Daffe. 1992. Lack of correlation between colony morphology and lipooligosaccharide content in the Mycobacterium tuberculosis complex. J. Gen. Microbiol. 138:1535–1541.

Mack D, Nedelmann M, Krokotsch A, Schwarzkopf A, Heesemann J, Laufs R. 1994. Characterization of transposon mutants of biofilm-producing *Staphylococcus epidermidis* impaired in the accumulative phase of biofilm production: genetic identification of a hexosamine containing polysaccharide intracellular adhesin. Infect. Immun. 62(8):3244–53

Minnikin, D. E., L. Kremer, L. G. Dover, and G. S. Besra. 2002. The methyl-branched fortifications of Mycobacterium tuberculosis. Chem. Biol. 9:545–553.

O'Toole GA, Kolter R. 1998. The initiation of biofilm formation in *Pseudomonas aeruginosa* WCS365 proceeds via multiple, convergent signaling pathways: a genetic analysis. *Mol. Microbiol.* 28:449

Ren, H., Dover, L.G., Islam, S.T., Alexander, D.C., Chen, J.M., Besra, G.S., Liu, J. (2007) Identification of the lipooligosaccharide biosynthetic gene cluster from Mycobacterium marinum. Mol. Microbiol. 63:1345–1359

Stewart, P.S. and Costerton, J.W. 2001. Antibiotic resistance of bacteria in biofilms. Lancet 358, 135–138

Wei, J., J. L. Dahl, J. W. Moulder, E. A. Roberts, P. O'Gaora, D. B. Young, and R. L.

Freidman. 2000. Identification of a *Mycobacterium tuberculosis* gene that enhances mycobacterial survival in macrophages. J. Bacteriol. 182:377–384.

Wingender et al. 1999. Microbial Extracellular Polymeric Substances. Springer Publsihing.

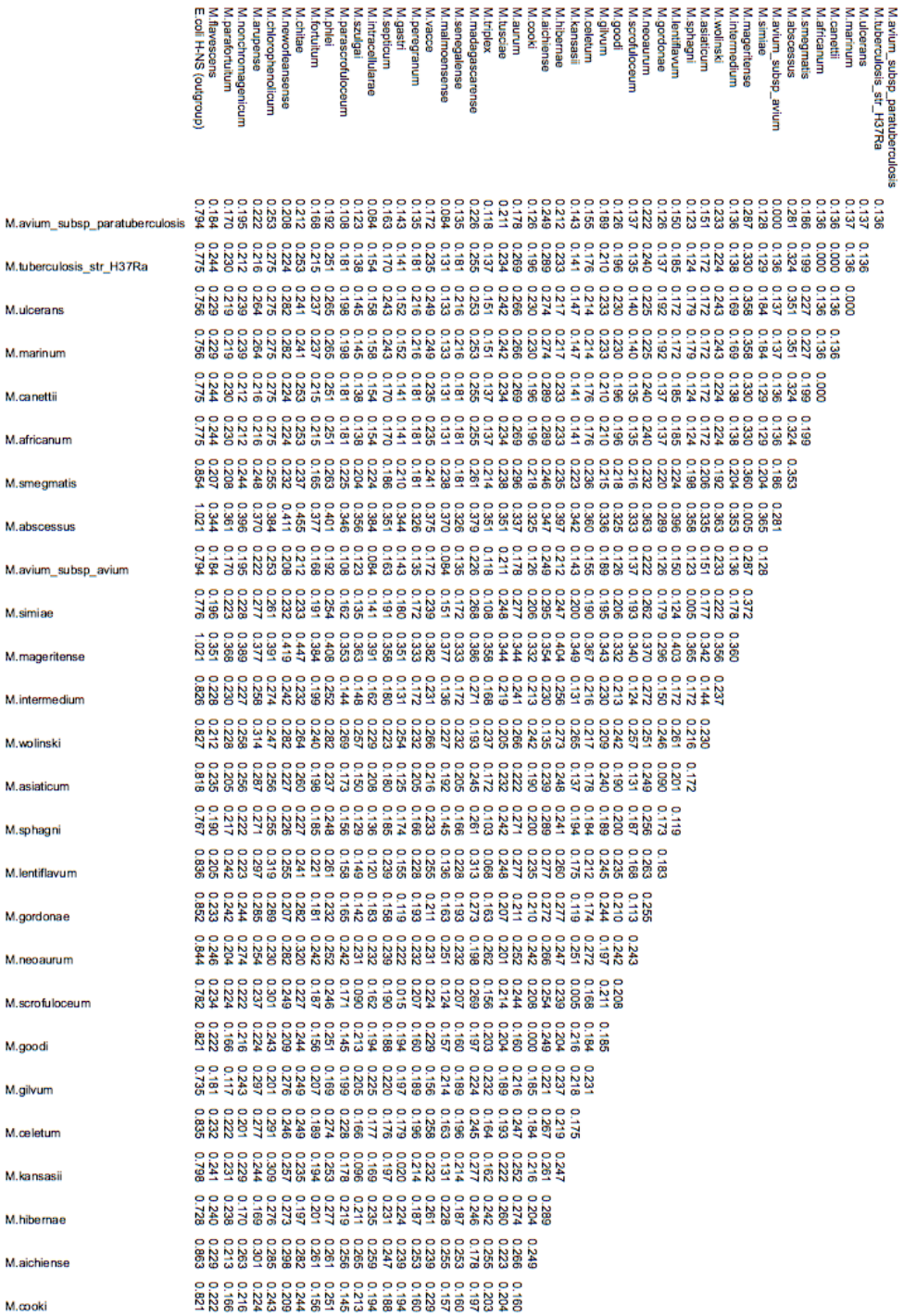
Table 1: List of *mycobacterial* strains and GenBank sequences

Bacterial Species	Reference strain
Mycobacterium aichiense	BAA495
Mycobacterium cookii	49103
Mycobacterium neoaurum	25795
Mycobacterium aurum	23366
Mycobacterium diernhoferi	19340
Mycobacterium hiberniae	49874
Mycobacterium lentiflavum	51985
Mycobacterium tusciae	BAA-564
Mycobacterium triplex	700071
Mycobacterium goodii	BAA-955
Mycobacterium sphagni	33027
Mycobacterium alvei	51304
Mycobacterium chelonae	35752
Mycobacterium asiaticum	25276
Mycobacterium madagascariense	49865
Mycobacterium senegalense	35796
Mycobacterium malmoense	29571
Mycobacterium vaccae	15483
Mycobacterium peregrinum	14467
Mycobacterium palustre	BAA-377
Mycobacterium gastri	15759
Mycobacterium septicum	700731
Mycobacterium intracellulare	13950
Mycobacterium simiae	25275
Mycobacterium gilvum	43909
Mycobacterium mageritense	700351
Mycobacterium szulgai	35799
Mycobacterium parascrofulaceum	BAA-614
Mycobacterium phlei	11758
Mycobacterium chubuense	27278
Mycobacterium fortuitum	6841
Mycobacterium porcinum	33776
Mycobacterium chitae	19628
Mycobacterium neworleansense	49404
Mycobacterium fortuitum	59403
Mycobacterium scrofulaceum	19981
Mycobacterium intermedium	51848
Mycobacterium wolinskyi	700010
Mycobacterium	49826

chlorophenolicum	
Mycobacterium arupense	BAA-124
Mycobacterium smegmatis	700084
Mycobacterium marinum	927
Mycobacterium gordonae	35756
Mycobacterium nonchromogenicum	19531
Mycobacterium abscessus	19977
Mycobacterium avium subsp. avium	25291
Mycobacterium triviale	23292
Mycobacterium mucogenicum	49649
Mycobacterium parafortuitum	25807
Mycobacterium kansasii	12478
Mycobacterium flavescens	14474
Mycobacterium celatum	51131
Mycobacterium tuberculosis H37Ra	NC_009525.1*
Mycobacterium tuberculosis H37Rv	NC_000962.3*
Mycobacterium ulcerans	NC_008611.1*
Mycobacterium bovis	NC_012207.1*
Mycobacterium leprae	NC_002677.1*
Mycobacterium canettii	NC_015848.1*
Mycobacterium africanum	NC_015758.1*
Mycobacterium avium subsp. paratuberculosis	NC_002944.2*

* indicates reference sequence was taken from NCBI database

Table 2: Interspecies differences of internal 250 bp *lsr2* gene sequence fragments. Numbers at each intersection of species represent the percentage of nucleotide differences between each pair.



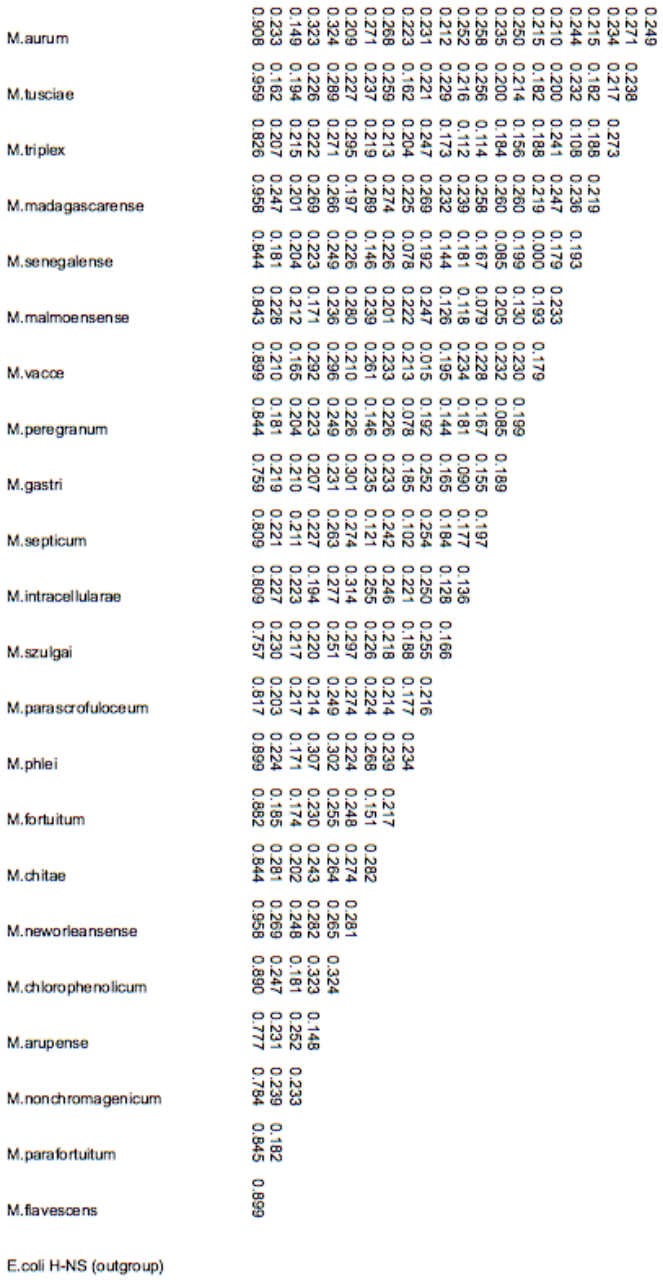

M.aurum
M.tusciae
M.triplex
M.madagascarense
M.senegalense
M.malmoensense
M.vacce
M.peregranum
M.gastri
M.septicum
M.intracellulare
M.szulgai
M.parascrofuloceum
M.phlei
M.fortuitum
M.chitae
M.neworleansense
M.chlorophenolicum
M.arupense
M.nonchromagenicum
M.parafortuitum
M.flavescens
E.coli H-NS (outgroup)

Table 3: List of strains and plasmids used in this study

Strain or Plasmid	Characteristic	Source
mc^2155	Wild-type strain	
Δlsr2	M. smegmatis with unmarked lsr2 mutation	(Arora, 2008)
Δlsr2Δ4727	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4727; Kan^r	This study
Δlsr2Δ4728	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4728; Kan^r	This study
Δlsr2Δ4729	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4729; Kan^r	This study
Δlsr2Δ4730	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4730; Kan^r	This study
Δlsr2Δ4731	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4731; Kan^r	This study
Δlsr2Δ4732	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4732; Kan^r	This study
Δlsr2Δ4733	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4733; Kan^r	This study
Δlsr2Δ4741	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_4741; Kan^r	This study
Δlsr2Δlsr2	M. smegmatis with unmarked lsr2 mutation;deactivated gene MSMEG_6092; Kan^r	This study
Plasmids		
pTOPO::'4727'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4727; Kan^r	This study
pTOPO::'4728'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4728; Kan^r	This study
pTOPO::'4729'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4729; Kan^r	This study
pTOPO::'4730'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4730; Kan^r	This study
pTOPO::'4731'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4731; Kan^r	This study
pTOPO::'4732'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4732; Kan^r	This study
pTOPO::'4733'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4733; Kan^r	This study
pTOPO::'4741'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_4741; Kan^r	This study
pTOPO::'6092'	Mycobacterial suicide vector carrying an internal fragment of MSMEG_6092; Kan^r	This study
pLSR2	lsr2 gene cloned into shuttle vector pNBV1;Kan^r	(Arora, 2008)

Table 4: PCR primers and templates used in this study

Primer N	Primer Sequence (5'->3')	Target Gene
4727' F	CGGGATTTTCACCGCAACTT	Small internal fragment of 4727
4727' R	GTTGCGGTTCTCCAGGAACT	Small internal fragment of 4727
4728' F	CTACCACAGCCGTTTCGAGT	Small internal fragment of 4728
4728' R	CAAGGAAGGTCACCCACTCC	Small internal fragment of 4728
4729' F	CGGGATTTTCACCGCAACTT	Small internal fragment of 4729
4729' R	GTTGCGGTTCTCCAGGAACT	Small internal fragment of 4729
4730' F	CTCTGGATCTACGAGCACGC	Small internal fragment of 4730
4730' R	GAGTCTCGTCAGTGGTCGTC	Small internal fragment of 4730
4731' F	TCGCACATCGCCGAATATCT	Small internal fragment of 4731
4731' R	TCACGTCGGGATTCTGGTTC	Small internal fragment of 4731
4732' F	GTGAAATTCGTGCGCAGTCC	Small internal fragment of 4732
4732' R	GTATCCACCCAGTGATCGCA	Small internal fragment of 4732
4733' F	GTCATCCCACCGCTTCAGAT	Small internal fragment of 4733
4733' R	AGGACACCGAACAATCCGAG	Small internal fragment of 4733
4741' F	GCACAGCAACATGGGTTACG	Small internal fragment of 4741
4741' R	TTGCGCATCTCGTTCTGGAT	Small internal fragment of 4741
6092' F	ATGGCKAAGAAAGTNACCGTCAC	Small internal fragment of 6092
6092' R	ATGACGTCGGCCGGGATSCGGCC	Small internal fragment of 6092
4734' F	ATCGAATCGGGTGACGAGTG	Small internal fragment of 4734
4734' R	CGAGGCTTCTTTGGTGAGGT	Small internal fragment of 4734
4735' F	CAGTCCTGTGACGACCATCTC	Small internal fragment of 4735
4735' R	GAACGACGTGCCGCAATAAC	Small internal fragment of 4735
4736' F	CGGCCTGAAACTGACAGGTA	Small internal fragment of 4736
4736' R	ACGATCCGCACCGTTCATC	Small internal fragment of 4736
4737' F	GGGTTTGTGGATCTTCCGGT	Small internal fragment of 4737
4737' R	CCAAAATCCCCGCATGCAAA	Small internal fragment of 4737
4738' F	GCCGAACCTCGCAACATTTA	Small internal fragment of 4738
4738' R	GACAAGTGCAGGCAGATCCT	Small internal fragment of 4738
4739' F	ACCGCACAAAGACCGGATAG	Small internal fragment of 4739
4739' R	ATCTCCCACTGCTGCATGAC	Small internal fragment of 4739
4740' F	GTTCGTGGAACGTATCGGGT	Small internal fragment of 4740
4740' R	CGTGATCGCTCTCGGGTATC	Small internal fragment of 4740
4727 F	GTCATCCCACCGCTTCAGAT	Large internal fragment of 4727
4727 R	AGGACACCGAACAATCCGAG	Large internal fragment of 4727
4728 F	TCCTGGCAAACATGTGTCGT	Large internal fragment of 4728
4728 R	CCCATCTCGCTGAAGATCCG	Large internal fragment of 4728
4729 F	TACGAACGTCCGGTCGAGA	Large internal fragment of 4729
4729 R	GCAGTTGTGGCTTGTCGTTC	Large internal fragment of 4729

4730 F	ACAACCGGCTGTCGTTCATC	Large internal fragment of 4730
4730 R	CTGGGCATCGAGCTGATAGG	Large internal fragment of 4730
4731 F	TCGCACATCGCCGAATATCT	Large internal fragment of 4731
4731 R	TCACGTCGGGATTCTGGTTC	Large internal fragment of 4731
4732 F	GTGAAATTCGTGCGCAGTCC	Large internal fragment of 4732
4732 R	GTATCCACCCAGTGATCGCA	Large internal fragment of 4732
4733 F	GTCATCCCACCGCTTCAGAT	Large internal fragment of 4733
4733 R	AGGACACCGAACAATCCGAG	Large internal fragment of 4733
4741 F	GCACAGCAACATGGGTTACG	Large internal fragment of 4741
4741 R	TTGCGCATCTCGTTCTGGAT	Large internal fragment of 4741

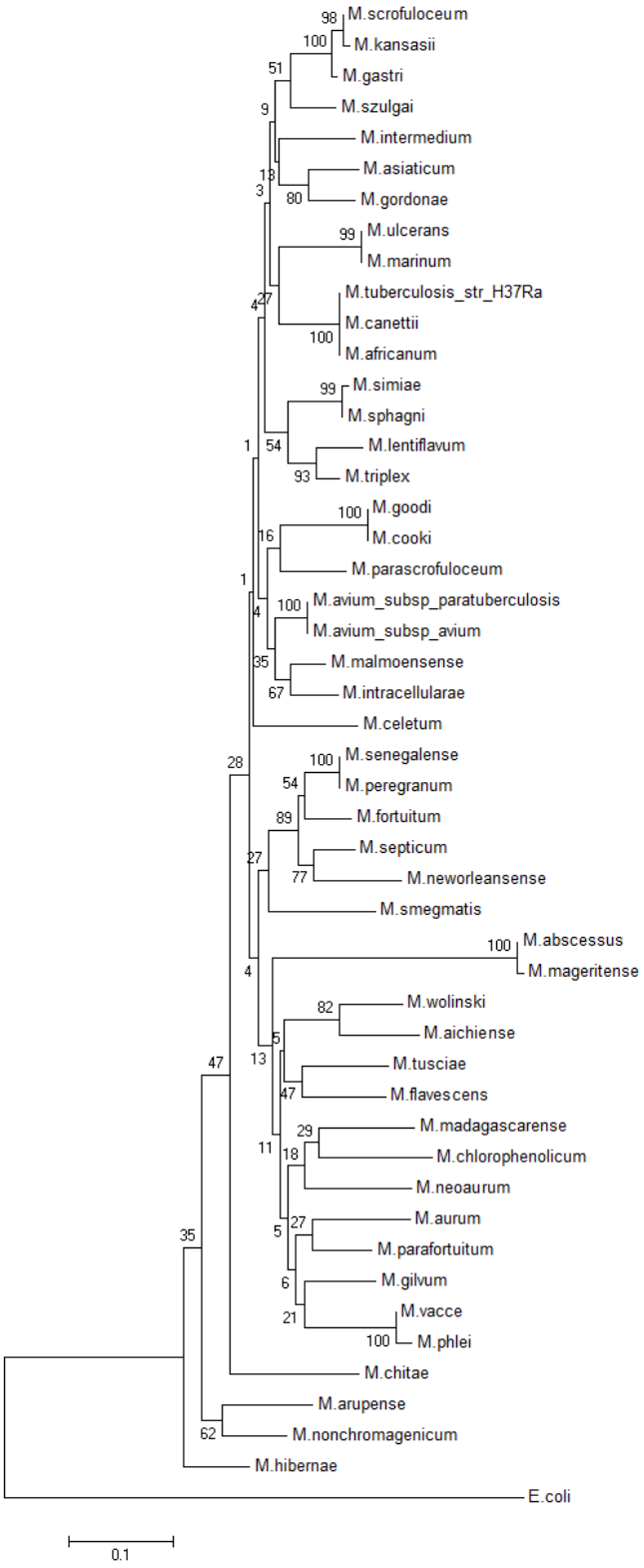


Figure 1A: Phylogenetic tree derived from *lsr2* nucleotide sequences. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1978). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstien 1987). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et. al. 2004) and are in the units of the number of base substitutions per site. The analysis involved 49 nucleotide sequences. Evolutionary analyses were conducted in MEGA5 (Tamura et. al. 2011).

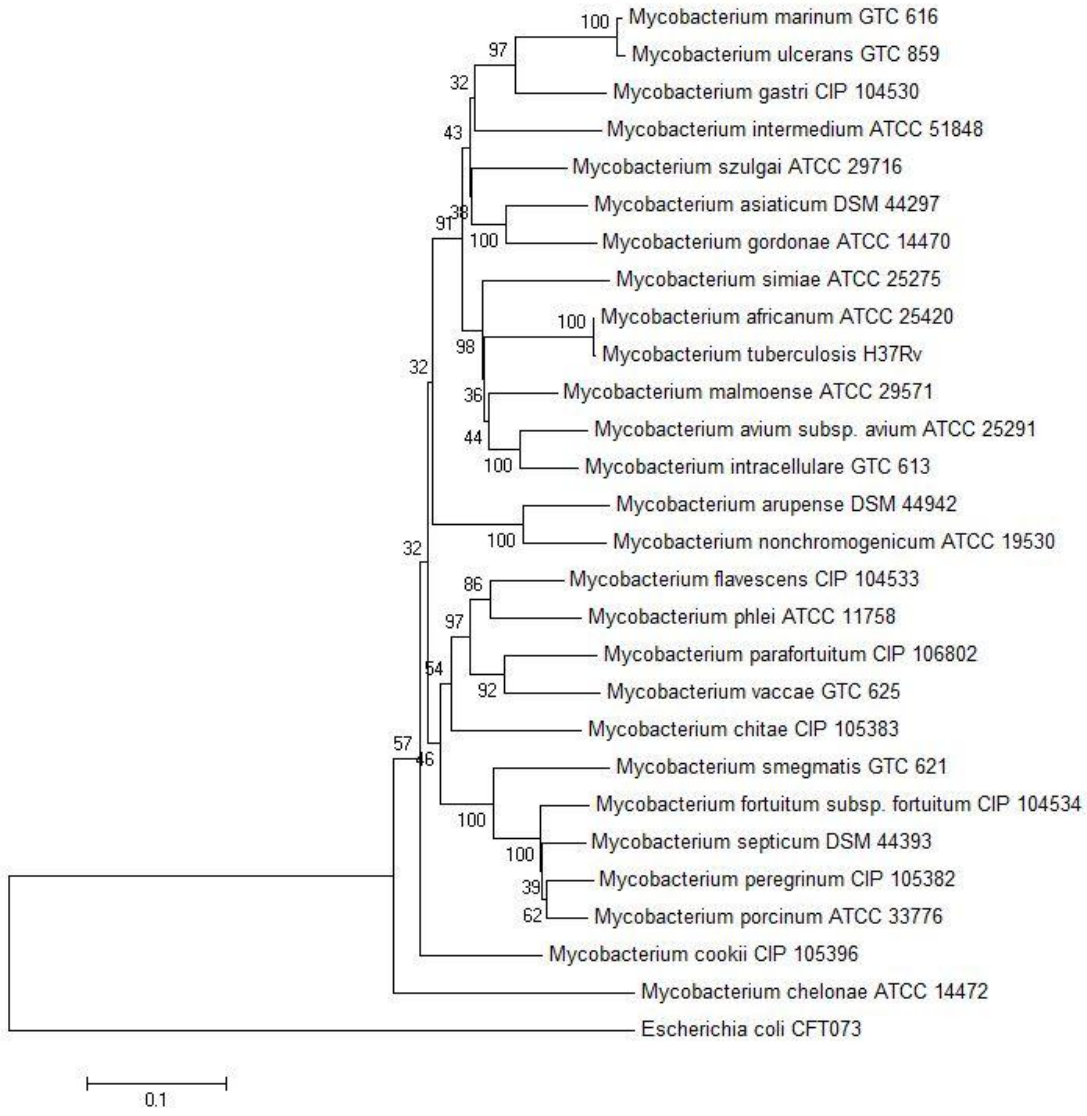


Figure 1B: Concatamer Phylogenetic Tree using dnaJ and lsr2

A 350-bp fragment of *dnaJ* was used with *lsr*2 to create a concatamer tree of the two sequences.

The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 2.77471284 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method [3] and are in the units of the number of base substitutions per site. The analysis involved 28 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1458 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

Figure 2:Multiple species alignment of four *M. smegmatis lsr2* sequences. The purpose of this figure is to show that there is no apparent difference when sequencing *lsr2* from PCR or plasmid constructs, and that high-fidelity *Taq* polymerase is not necessary. The top line of the MSA is a *lsr2* sequence from GenBank. Line 2 is the result of sequencing pTOPO with *lsr2* ligated into the multiple cloning site. Line 3 is a sequenced PCR product, as is line 4. The PCR product in line 3 was produced using the method described in the methods section, while the PCR product in line 4 used high fidelity *Taq* polymerase. All sequences are identical.

Figure 3: Alignment of internal *lsr2* sequence fragments from sequencing reactions and Genbank sequences. Clustal color scheme was used to show nucleotide differences and *E. coli* was used as an outgroup. PCR primer regions have been trimmed from the sequences to prevent false alignment. The sequences start at position 69 and end at position 323 in *M. smegmatis*. The EcoRI restriction site can be seen from positions 1-6 of the alignment, and represents the most well conserved run of 6 consecutive nucleotides in the alignment. The bar graph at the base of the graphic shows a visual consensus at each nucleotide.

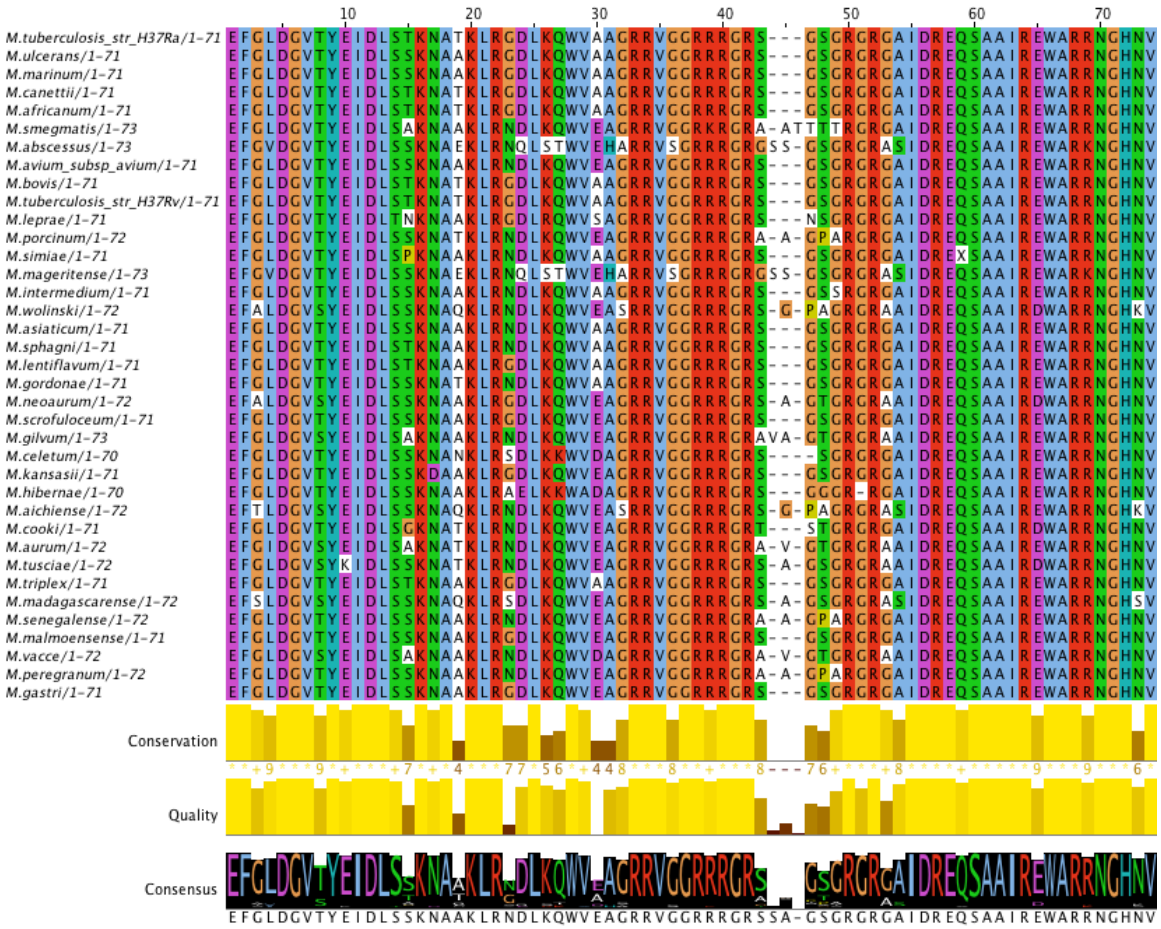


Figure 4: Alignment of translated Lsr2 amino acid sequence fragments.

The conservation and consensus of each alignment position is indicated by the bars at the base of the graphic. The conservation bar visually indicates how conserved each amino acid was at the indicated position. High conservation is represented by an asterisk, and lower conservation is represented by the height of the bar and a number from 1-9 below the bar. Amino acid consensus is also shown graphically, with the relative size of the single-letter amino acid abbreviation representing it's prevalence at each position. The most prevalent residue is indicated below each bar. Clustal color scheme was used to show amino acid differences in the graphic for quick visual analysis.

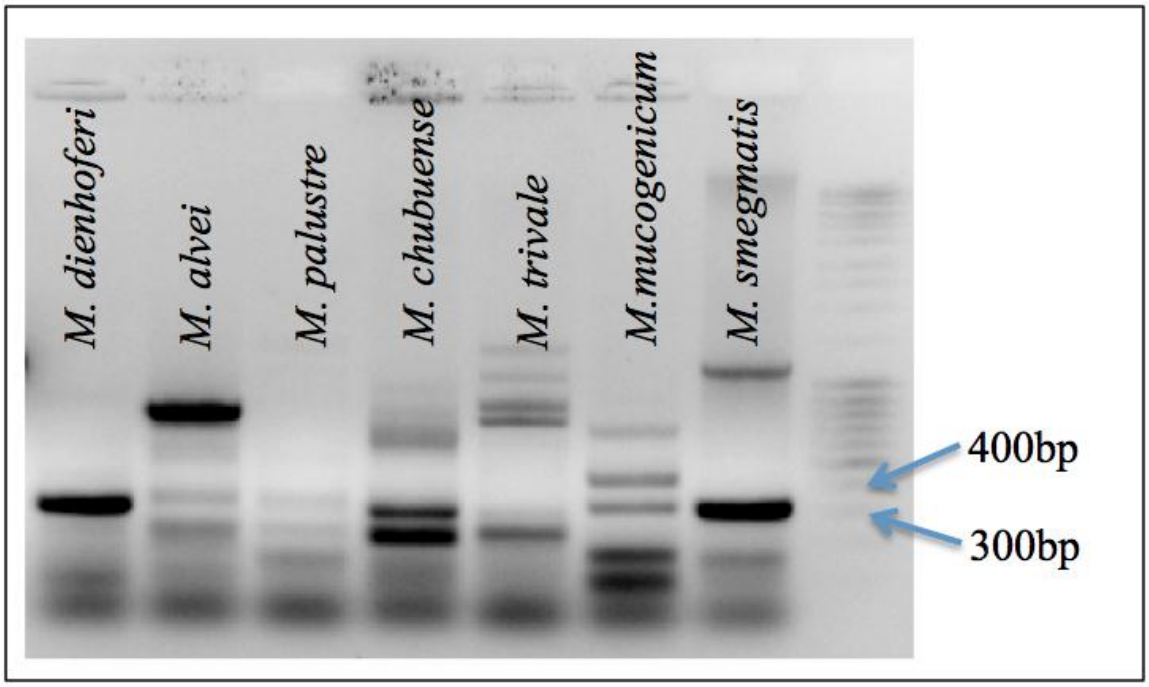


Figure 5: Agarose gel showing PCR products from 6 mycobacterial species that tested negative for *lsr2* and *M. smegmatis* as a control. Arrows indicate the 400 and 300bp weights on the MassRuler DNA ladder (Thermo). The 325bp *lsr2* fragment can be seen in *M. smegmatis*, but is less clear in each of the other 6 species tested. Sequencing of the band seen in *M. dienhoferi* yielded a product that did not appear to be *lsr2* when analyzed. It is possible that *lsr2* exists in these species, but due to non-specific binding of PCR primers it was not possible to isolate the fragment.

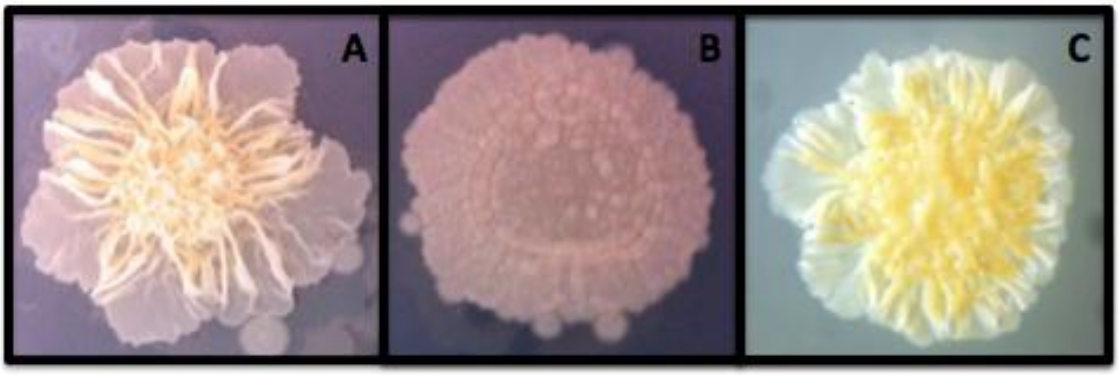


Figure 6: Colony morphologies of *M. smegmatis*. Each image is of one bacterial colony of *M. smegmatis* growing on 7H11 agar and magnified with a dissection microscope. Box A shows a WT colony with rough edges and wrinkled interior. Box B shows the Δ*lsr2* phenotype, which has smooth edges and colony interior. Box C shows the Δ*lsr2*Δ*mas* colony, which apprears more like the WT colony than the Δ*lsr2* colony.

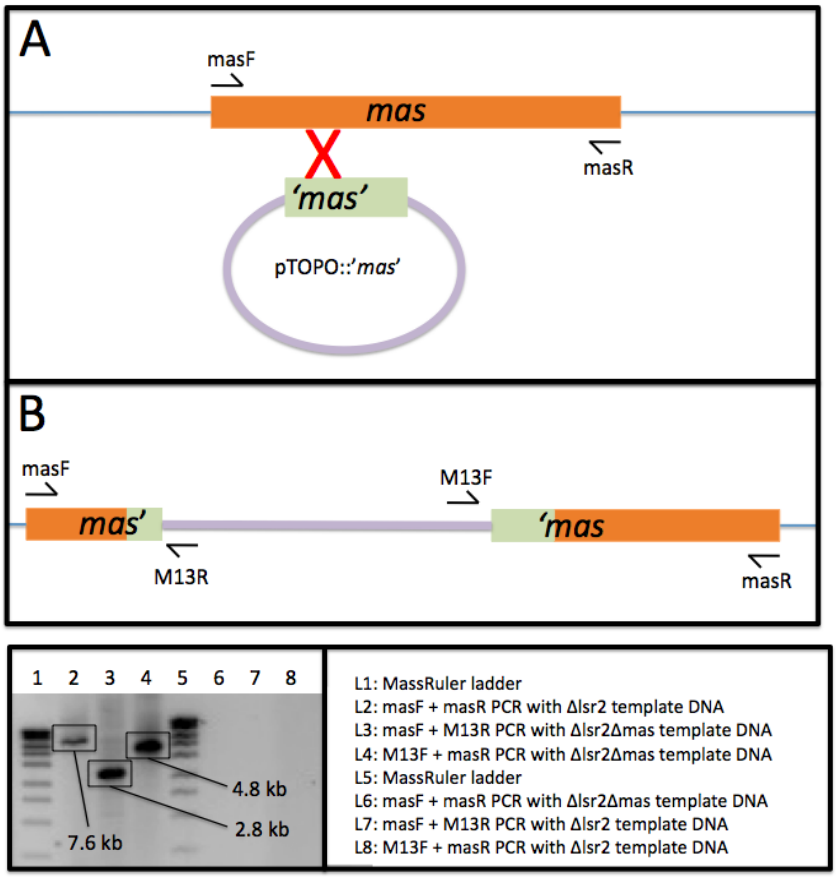


Figure 7: Diagram of the meridiploid knockout of *mas* via homologous recombination. In this figure **Box A** shows the manner in which homologous recombination would take place between the native chromosomal *mas* gene and the truncated PCR fragment *'mas'* of pTOPO::*'mas'*. The X marked an example area where homologous recombination could have taken place, although the exact location was not known.

Box B illustrates the insertion of the plasmid into chromosomal *mas.* The original chromosomal *mas* was broken into two incomplete fragments, the first (mas') missing the 5' end of the gene, and the second ('mas) missing the 3' end.

An agarose gel containing PCR fragments used to verify homologous recombination can be seen in **Box C**. The template for PCR in lanes 2-4 was chromosomal *mas* that had undergone homologous recombination and the template in lanes 6-8 was WT *mas.* The band visible in lane 2 was a 7.6 Kb PCR fragment generated by amplifying *mas* in *M. smegmatis* Δ*lsr2* with primer masF and masR (arrows in box A). When the same primer pair is used on *M. smegmatis* Δ*lsr2*Δ*mas* the expected band size is 12.8 Kb but no band is observed in lane 6 likely because of the large amplification size (arrows in box B). Lanes 3 and 4 show the PCR product size obtained when pTOPO M13F and M13R primers were used in conjunction with masF and masR (arrows in box B). No bands were seen in lanes 7 and 8 as the M13 primers were unable to bind in the absence of a pTOPO backbone.

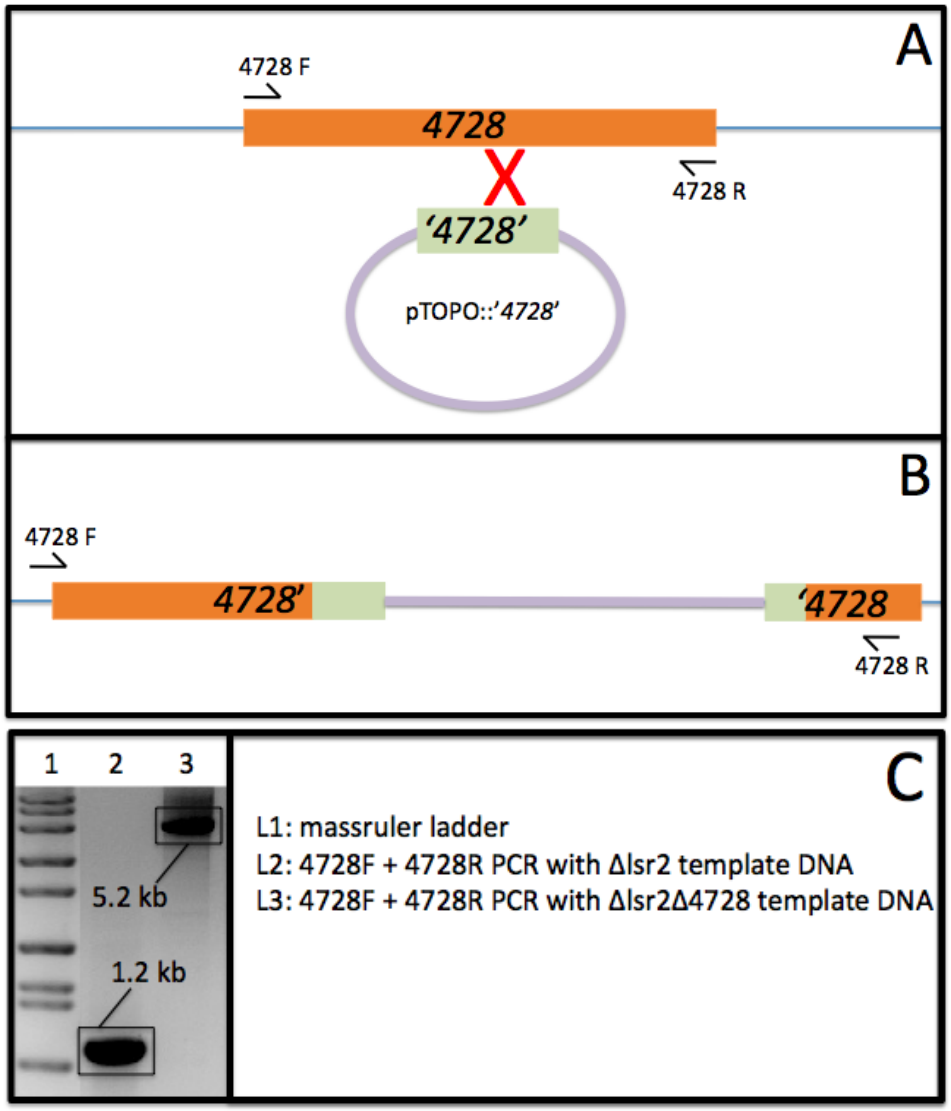


Figure 8: Diagram of the meridiploid knockout of 4728 via homologous recombination. In this figure **Box A** shows the manner in which homologous recombination would take place between the native chromosomal *4728* gene and the truncated PCR fragment *'4728'* of pTOPO::*'4728'*. The X marked an example area where homologous recombination could have taken place, although the exact location was not known.

Box B illustrates the insertion of the plasmid into chromosomal *4728*. The original chromosomal *4728* was broken into two incomplete fragments, the first (4728') missing the 5' end of the gene, and the second ('4728) missing the 3' end.

An agarose gel containing PCR fragments used to verify homologous within 4728 can be seen in **Box C**. Lane 1 was massruler DNA ladder. Lane 2 was the PCR product created when primers 4728F and 4728R were used with Δlsr2 template DNA. The PCR product seen in lane 3 was formed using the same primer pair on the Δlsr2Δ4728 template DNA. The band size increase was due to insertion of the pTOPO backbone.

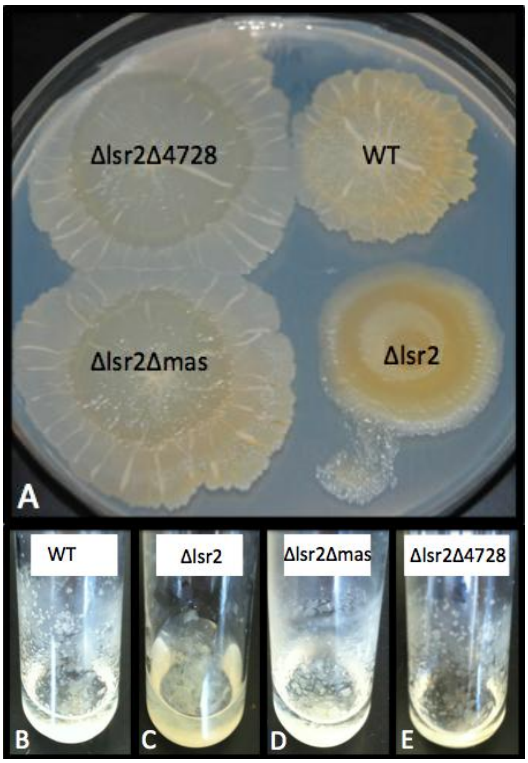


Figure 9: Colony morphology and biofilm formation assays for *M. smegmatis* mutant strains. **Box A** showed the result of spotting *M. smegmatis* mutant and WT strains on motility agar. The WT *M. smegmatis* has a rough texture to the colony with ridges as the spot expanded outward while the Δlsr2 *M. smegmatis* displays a smooth colony morphology. The Δlsr2Δmas and Δlsr2Δ4728 colonies displayed increased ability to spread over the soft agar and had a rough texture that resembled the WT.

Box B shows WT *M. smegmatis* growing in a test tube with containing 7H11 medium. After 1 week of growth a biofilm was observed growing up the sides of the test tube, and clumps of cells gathered at the bottom. A pellicle was observed floating at the ALI.

Box C shows the Δlsr2 *M. smegmatis* growing in a test tube with containing 7H11 medium. After 1 week of growth no biofilm or pellicle was observed, and cells were suspended evenly in the growth media.

Box D and **Box E** shows Δlsr2Δmas and Δlsr2Δ4728 *M. smegmatis*, respectively, growing in a test tube with containing 7H11 medium. After 1 week of growth a biofilm was observed growing up the sides of both test tubes, and clumps of cells gathered at the bottom. A pellicle was observed floating at the ALI in both tubes.

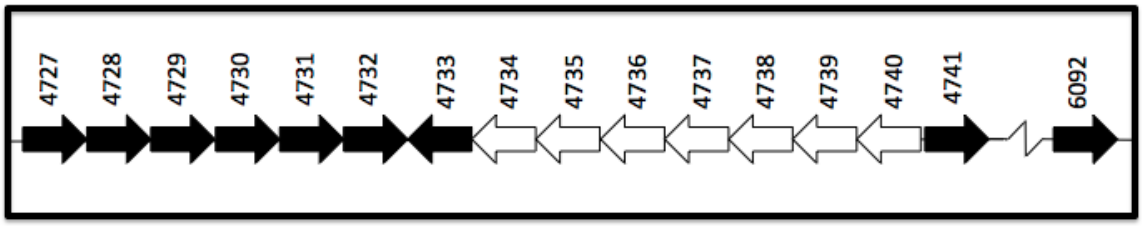


Figure 10: Cartoon diagram of genes examined in this work. The "mas cluster" begins with 4727 (*mas*) and ends at or before 4741. Black arrows depict genes that were inactivated during this study. The *lsr2* gene (6092) is distant to the "mas cluster", but was also inactivated as a control. Genes depicted by white arrows were not assessed in this study.

Comprehensive Bibliography

Arora, K., D. C. Whiteford, D. Lau-Bonilla, C. M. Davitt, and J. L. Dahl. 2008. Inactivation of *lsr2* results in a hypermotile phenotype in *Mycobacterium smegmatis*. J. Bacteriol. 190:4291–4300.

Branda SS, Vik S, Friedman L, Kolter R. 2005. Biofilms: the matrix revisited. Trends Microbiol.;13:20–26.

Chen, J., German, G., Alexander, D., Ren, H., Tan, T., and Liu, J. 2005. Roles of Lsr2 in Colony Morphology and Biofilm Formation of Mycobaterium smegmatis. Journal of Bacteriology. 633-641.

Chen, J. M., H. Ren, J. E. Shaw, Y. J. Wang, M. Li, A. S. Leung, V. Tran, N. M. Berbenetz, D. Kocincova, C. M. Yip, J. M. Reyrat, and J. Liu. 2008. Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. Nucleic Acids Res. 36:2123–2135.

Colangeli, R. ,A. Haq, V. L. Arcus, E. Summers, R. S. Magliozzo, A. McBride, A. K. Mitra, M. Radjainia, A. Khajo, W. R. Jacobs, Jr., P. Salgame, and D. Alland. 2009. The multifunctional histone-like protein Lsr2 protects *mycobacteria* against reactive oxygen intermediates. Proc. Natl. Acad. Sci. U. S. A. 106:4414–4418.

Colangeli, R., D. Helb, C. Vilcheze, M. H. Hazbon, C. G. Lee, H. Safi, B. Sayers, I. Sardone, M. B. Jones, R. D. Fleischmann, S. N. Peterson, W. R. Jacobs, Jr., and D. Alland. 2007. Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in *M. tuberculosis*.

Etienne G., Malaga W., Laval F., Lemassu A., Guilhot C., Daffé M. 2009. Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. J. Bacteriol. 191, 2613– 2621

Falkinham JO 3rd, Norton CD, LeCheva- llier MW. 2001. Factors influencing numbers of *Mycobacterium avium, Mycobacterium intracellulare*, and other *Mycobacteria* in drinking water distribution systems. Appl. Environ. Microbiol. 67:1225–31

Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. 2009. Opportunistic pathogens enriched in showerhead biofilms. Proceedings of the National Academy of Sciences U S A.;106:16393–9

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

Geesey GG, Richardson WT, Yeomans HG, Irvin RT, Costerton JW. 1977. Microscopic examination of natural sessile bacterial populations from an alpine stream. *Can. J. Microbiol.* 23(12):1733–36

Gordon BR, Imperial R, Wang L, Navarre WW, Liu J. 2008. Lsr2 of Mycobacterium represents a novel class of H-NS-like proteins. Journal of bacteriology 190: 7052–7059

Gordon, B. R., Y. Li, L. Wang, A. Sintsova, H. van Bakel, S. Tian, W. W. Navarre, B. Xia, and J. Liu. 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. 107:5154–5159.

Hall-Stoodley, L., and H. Lappin-Scott. 1998. Biofilm formation by the rapidly growing *mycobacterial* species Mycobacterium fortuitum. FEMS Microbiol. Lett. 168:77–84.

Hall-Stoodley, L., J. W. Costerton, and P. Stoodley. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat. Rev. Microbiol. 2:95–108.

Hall-Stoodley, L., Brun, O.S., Polshyna, G., and Barker, L.P. 2006 *Mycobacterium marinum* biofilm formation reveals cording morphology. FEMS Microbiol Lett 257: 43–49.

Heilmann C, Gerke C, Perdreau-Remington F, Gotz F. 1996. Characterization of Tn917 insertion mutants of *Staphylococcus epidermidis* affected in biofilm formation. *Infect. Immun.* 64(1):277–82

Kim B. J., Lee S. H., Lyu M. A., Kim S. J., Bai G. H., Chae G. T., Kim E. C., Cha C. Y., Kook Y. H. 1999 Identification of *mycobacterial* species by comparative sequence analysis of the RNA polymerase gene (rpoB). J. Clin. Microbiol. 37:1714–1720.

Laal, S., Y. D. Sharma, H. K. Prasad, A. Murtaza, S. Singh, S. Tangri, R. S. Misra, and I. Nath. 1991. Recombinant fusion protein identified by lepromatous sera mimics native *Mycobacterium leprae* in T-cell responses across the leprosy spectrum. Proc. Natl. Acad. Sci. U. S. A. 88:1054–1058.

Lemassu, A., V. V. Levy-Frebault, M.-A. Laneelle, and M. Daffe. 1992. Lack of correlation between colony morphology and lipooligosaccharide content in the Mycobacterium tuberculosis complex. J. Gen. Microbiol. 138:1535–1541.

Mack D, Nedelmann M, Krokotsch A, Schwarzkopf A, Heesemann J, Laufs R. 1994. Characterization of transposon mutants of biofilm-producing *Staphylococcus epidermidis* impaired in the accumulative phase of biofilm production: genetic identification of a hexosamine containing polysaccharide intracellular adhesin. Infect. Immun. 62(8):3244–53

Minnikin, D. E., L. Kremer, L. G. Dover, and G. S. Besra. 2002. The methyl-branched fortifications of Mycobacterium tuberculosis. Chem. Biol. 9:545–553.

Nguyen K. T., Piastro K., Gray T. A., Derbyshire K. M.. 2010. Mycobacterial biofilms facilitate horizontal DNA transfer between strains of Mycobacterium smegmatis. J. Bacteriol. 192:5134–5142.

O'Toole GA, Kolter R. 1998. The initiation of biofilm formation in *Pseudomonas aeruginosa* WCS365 proceeds via multiple, convergent signaling pathways: a genetic analysis. *Mol. Microbiol.* 28:449

Ojha, A. K., A. D. Baughn, D. Sambandan, T. Hsu, X. Trivelli, Y. Guerardel, A. Alahari, L. Kremer, W. R. Jacobs, Jr., and G. F. Hatfull. 2008. Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol. Microbiol. 69:164-174

Ren, H., Dover, L.G., Islam, S.T., Alexander, D.C., Chen, J.M., Besra, G.S., Liu, J. (2007) Identification of the lipooligosaccharide biosynthetic gene cluster from Mycobacterium marinum. Mol. Microbiol. 63:1345–1359

Ringuet, H., C. Akoua-Koffi, S. Honore, A. Varnerot, V. Vincent, P. Berche, J. L. Gaillard, and C. Pierre-Audigier. 1999. hsp65 sequencing for identification of rapidly growing mycobacteria. J. Clin. Microbiol. 37:852-857

Sassetti, C. M. & Rubin, E. J. 2003. Genetic requirements for *mycobacterial* survival during infection. Proc. Natl Acad. Sci. USA 100, 12989–12994

Saitou N. and Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.

Stewart, P.S. and Costerton, J.W. 2001. Antibiotic resistance of bacteria in biofilms. Lancet 358, 135–138

Stoodley, P., K. Sauer, D. G. Davies, and J. W. Costerton. 2002. Biofilms as complex differentiated communities. Annu. Rev. Microbiol. 56:187– 209.

Tamura K., Nei M., and Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (*USA*) **101**:11030-11035.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

Wei, J., J. L. Dahl, J. W. Moulder, E. A. Roberts, P. O'Gaora, D. B. Young, and R. L.

Freidman. 2000. Identification of a *Mycobacterium tuberculosis* gene that enhances mycobacterial survival in macrophages. J. Bacteriol. 182:377–384.

Wingender et al. 1999. Microbial Extracellular Polymeric Substances. Springer Publsihing.

Yamada-Noda, M., Ohkusu, K., Hata, H., Shah, M. M., Nhung, P. H., Sun, X. S., Hayashi, M. & Ezaki, T. 2007. Mycobacterium species identification – a new approach via dnaJ gene sequencing. Syst Appl Microbiol 30, 453–462.

Zelazny, A. M., Calhoun, L. B., Li, L., Shea, Y. R. & Fischer, S. H. 2005. Identification of Mycobacterium species by secA1 sequences. J Clin Microbiol 43, 1051–1058